

tanulmányok

86/1978



MTA Számítástechnikai és Automatizálási Kutató Intézet Budapest



MAGYAR TUDOMÁNYOS AKADÉMIA
SZÁMITÁSTECHNIKAI ÉS AUTOMATIZÁLÁSI KUTATÓ INTÉZETE

A SIS77 STATISZTIKAI INFORMÁCIÓS RENDSZER
KIALAKÍTÁSÁNAK SZEMPONTJAI, ALKALMAZÁSÁNAK
ÉS TOVÁBBFEJLESZTÉSÉNEK LEHETŐSÉGEI

Írta:

Ruda Mihály

A kiadásért felelős:

DR VÁMOS TIBOR

ISBN 963 311 073 4

ISSN 0324-2951

TARTALOMJEGYZÉK

<u>BEVEZETÉS</u>	5
<u>1. A MAGYARORSZÁGI KÓRHÁZI MORBIDITÁSI VIZSGÁLAT</u>	
<u>- A SIS77 ALKALMAZÁSI HÁTTERE</u>	7
1.1 A magyarországi kórházi morbiditási vizsgálatok	7
1.1.1 A kezdeti lépések	7
1.1.2 Az 1967. évi vizsgálat	7
1.2 Az 1972-73. évi kórházi morbiditási vizsgálat	8
1.2.1 A vizsgálat szervezése	8
1.2.2 A számítógépes feldolgozás	9
1.2.3 Problémák és eredmények	10
1.3 A kórházi morbiditási vizsgálatok célja és módszertana, szervezési és számítástechni- kai problémák	11
1.3.1 Célok	11
1.3.2 Módszertani kérdések	13
1.3.3 A számítógépes rendszer	15
1.4 A SIS77 kialakulásának folyamata	20
1.5 A SIS77 alkalmazásának továbbfejlesztési lehetőségei	23
<u>2. A SIS77 SZERKEZETE ÉS MŰKÖDÉSI ELVEI,</u> <u>FELHASZNÁLÁSI LEHETŐSÉGEI</u>	25
2.1 A statisztikai adatfeldolgozás egy modellje	26
2.1.1 Adatbázisok - statisztikai adat- feldolgozás	26
2.1.2 Egy modellvázlat	30
2.1.3 A modell megvalósítása	37
2.2 Optimalizálási lehetőségek	41
2.2.1 Az optimalizálás jelentősége	41
2.2.2 Néhány optimalizálási feladatkör ..	43

2.3 A SIS77 felépítése, alkalmazási lehetőségek	50
2.3.1 A rendszer célja, felépítése	50
2.3.2 Ellenőrzés, új adatok létrehozása ..	54
2.3.3 Mintakiválasztás, az adatrendszer felbontása	57
2.3.4 Az alapadatforma konverziója	59
2.3.5 Speciális file-ok	61
2.3.6 Táblafile-ok létrehozása	62
2.3.7 Az adatok /eredmények/ megjelenítése	67
2.3.8 A rendszeradminisztráció	71
2.3.9 Segédfile-ok	74
2.3.10 Kapcsolatok más rendszerekkel	74
3. <u>A SIS77 TOVÁBBFEJLESZTÉSI LEHETŐSÉGEI</u>	76
3.1 Az alkalmazás kiterjesztése	77
3.1.1 Adatelőkészítés, az adattartalom bő- vítése, áttérés más számítógépekre .	77
3.1.2 Az adatbázis továbbfejlesztése, interface-ek	80
3.1.3 Az input és output formátumok fejlesztése	82
3.2 A szervezés és a belső felépítés kérdései	84
3.2.1 A szervezést segítő kiegészítések ..	84
3.2.2 A rendszerműködés hatékonyságának növelése	85
I R O D A L O M J E G Y Z É K	89

B e v e z e t é s

Az MTA SZTAKI Valószínűségszámítási és Matematikai Statisztikai Osztályán /Információs és Irányítási Rendszerek Főosztály/ egy csoport már több éve foglalkozik egészségügyi statisztikai /és matematikai statisztikai/ feladatokkal. Ebben a témakörben központi szerepet játszik a kórházi morbiditási vizsgálat. Az ESZTIK /Egészségügyi Minisztérium Szervezési, Tervezési és Információs Központja/ és a SZTAKI 1976. novemberében kötött kutatási és fejlesztési szerződésének keretén belül az előbb említett csoport munkatársai, az országos kórházi morbiditási adatok feldolgozásának céljára létrehoztak egy általános statisztikai adatfeldolgozó rendszert. Az általános megoldásmódból eredően azonban nem csak bizonyos rögzített morbiditási statisztikák szolgáltatására alkalmas programokat készítettünk el, hanem egy olyan software-eszközt amely egyrészt a kórházi morbiditási rendszer változása esetén is használható, másrészt egyéb statisztikai felmérések feldolgozásánál is alkalmazható. Ezt az általános használhatóságot fejezi ki a rendszer elnevezése is: SIS77 - Statistical Information System, 1977.

A tanulmány első részében a SIS77 létrehozásának körülményeit, az országos kórházi morbiditási vizsgálatok alakulását tárgyalva áttekintést adunk a rendszer létrehozásának folyamatáról és a rendszer kialakulására ható tényezőkről.

A második részben a SIS77 szerkezeti és működési elveit mutatjuk be. Ez egyben módot ad arra, hogy a felhasználási lehetőségeket és más rendszerekhez való kapcsolódás formáit is leírjuk.

A SIS77 szerkezetéből adódóan könnyen továbbfejleszthető, kiegészíthető és más rendszerekhez is kapcsolható. A tanulmány harmadik részében ezeket a lehetőségeket vizsgáljuk meg.

A szerző köszönetét fejezi ki ARATÓ MÁTYÁSNAK a rendszermegvalósítás munkáinak támogatásáért és értékes szakmai tanácsaiért. Ugyancsak köszönet illeti KRÁMLI ANDRÁST, aki mint a téma egyik felelőse résztvett a rendszerkoncepció kialakításában.

Köszönettel tartozunk az Egészségügyi Minisztérium és az ESZTIK vezetőinek és munkatársainak, megtisztelő bizalmukért és - munkánkat jelentős mértékben segítve - a rendszer alkalmazásában történő aktív részvételükért.

1. A magyarországi kórházi morbiditási vizsgálat - a SIS77 alkalmazási háttere

1.1 A magyarországi kórházi morbiditási vizsgálatok

1.1.1 A kezdeti lépések. Magyarországon kórházi adatgyűjtés már a múlt század 70-es éveitől kezdve folyik. A kórházi betegek adatainak részletesebb vizsgálatára azonban csak az ötvenes évek elején tették meg az első lépést. Egységes statisztikai adatszolgáltatás 1953 óta van kórházainkban, és ezen az alapon 1955 és 1965 között minden évben végeztek statisztikai felméréseket a hospitalizált betegekről. Ezeket a felméréseket azonban kézi uton dolgozták fel, így természetesen csak a legegyszerűbb statisztikákat nyerhették.

1.1.2 Az 1967. évi vizsgálat. Az előző vizsgálatoknál lényegesen több lehetőséget nyújtott az 1967-es felmérés, amelyet már gépi uton végeztek, és így összetettebb kérdésekre is feleletet lehetett kapni. A felhasznált gépi eszköz viszont még mindig nem egy korszerű, nagyteljesítményű elektronikus számológép

volt, hanem Hollerith gépek. Ez lelassította és bonyolulttá tette a feldolgozást.

Tanulmányunknak nem célja a kórházi morbiditási vizsgálatok részletes bemutatása. Az eddigi eredményekről számos publikáció jelent meg az Egészségügyi Minisztérium és az ESZTIK kiadványaiban, a Statisztikai Szemlében, a Népegészségügyben, az Egészségügyi Gazdasági Szemlében. Ezek közül említjük a [8] és [25] tanulmányokat. A kórházi morbiditási vizsgálatok egy rövid történeti áttekintését adja [18].

Részletesebben foglalkozunk az 1972-73. évi kórházi morbiditási vizsgálattal, amely az első valóban korszerű eszközöket alkalmazó felmérés volt, és kiinduló pontja volt a SIS77 megvalósításának is.

1.2 Az 1972-73. évi kórházi morbiditási vizsgálat

1.2.1 A vizsgálat szervezése. Az előző pontban vázolt kezdeti lépések után került sor az első számítógépes adatfeldolgozásra. Ennél a munkánál tehát alapvető számítástechnikai tapasztalatokat is kellett szerezni a hospitalizált morbiditási rendszer vonatkozásában. Az 1972-73. évben 14 hónapon keresztül rögzítették a kórházakból kiírt betegek 36 százalékának adatait

/a mintavétel szakmánként változó, 33 illetve 50 százalékos volt/. Így több mint 700 ezer ápolási eset került feldolgozásra /ebből az első két hónap - mintegy 100 ezer eset - próbafelvétel volt/. A szakemberek a különböző szintű igények alapján 37 statisztikai táblázat tervét készítették el. Bár ezek a táblatervek a feldolgozás folyamán nagyrészt módosultak és jelentős kiegészítések is történtek, mégis az egész vizsgálat alapját ezek az előre rögzített táblázatok képezték /erre a kérdéskörre még visszatérünk/. A felmérés eredményeit a [18] és [19] kötet tartalmazza.

1.2.2 A számítógépes feldolgozás. A felmérés számítógépes feldolgozását az MTA SZTAKI Valószínűségsszámítási és Matematikai Statisztikai Osztályának munkatársai végezték el /az MTA CDC 330-as gépén - ld. [26]/. A feladat számos érdekes számítástechnikai és matematikai problémát vetett fel, hiszen nemcsak nagytömegű - mintegy 70-80 millió karakternyi - adat feldolgozásáról volt szó, hanem az elkészítendő statisztikai táblázatok is egy bonyolult rendszert alkottak /a megbízók a hospitalizált betegekről, a kórházak működéséről széleskörű és mélyreható információkat kívántak nyerni/, és a bonyolultságon túl a

megjelenítésre kerülő adatok mennyisége is igen nagy volt. Ez utóbbi kérdésben példaként szolgálhat a [19] kötet, amely tulajdonképpen egyetlen táblázatot közül 280 oldalon keresztül /az előbb említett 37 táblázat közül/. Meg kell jegyezni azt is, hogy ezt a feladatot a ma rendelkezésre álló /ÁSZSZ HWB 66/60 /gépkapacitásnál lényegesen kisebb erőforrások felhasználásával /az Akadémia azóta már jelentős mértékben bővített CDC 3300-as gépén/ kellett megoldani.

1.2.3 Problémák és eredmények. A felmerülő problémák természetesen sok értékes kutatási feladatot jelentettek, melyek eredményeként általánosan is használható software-eszközökhöz jutottunk /ld. pl. [11] / - erre a későbbiekben még részletesebben visszatérünk. Eredményeinkről és a feldolgozás menetéről az Egészségügyi Minisztérium munkatársaival együtt több publikációban beszámoltunk /ld. [9], [10], [11], [15], [17], [26]/. A feldolgozási munkák lehetővé tették, hogy a kórházi morbiditási vizsgálat témakörében jelentős mértékben tájékozódjunk. Ezeket a tapasztalatokat a későbbiek folyamán jól hasznosítottuk.

Mindemellett az 1972-73. évi felmérés egy kísérleti jellegű vizsgálat volt. Tapasztalatokat kellett szerezni nagytömegű morbiditási adat számítógépes feldolgozásában, és a lehetséges igényeket, a meg-

valósítási lehetőségeket is fel kellett mérni. Ez egyrészt orvosi, szervezési, másrészt számítástechnikai, matematikai feladatok megoldását igényelte.

1.3 A kórházi morbiditási vizsgálatok célja és módszertana, szervezési és számítástechnikai problémák

1.3.1 Célok. A korszerű egészségügyi ellátás egyik nélkülözhetetlen alapja az ellátásról és az igényekről nyerhető információk gyűjtése, és az érdekeltek /vezetési, szervezési és gyógyító intézetek illetve személyek/ felé való továbbítása. Az egészségügyi információs rendszer egyik leglényegesebb eleme a hospitalizált betegek és a fekvőbeteggyógyító intézetek adatait hordozó részrendszer. A kórházi morbiditási vizsgálat az előző pontokban bemutatott felmérések folyamán egyre több és részletesebb információt szolgáltatott. A mai igényeket nagyjából a következőkben lehet összefoglalni. /Részletesebben ld. pl. [9], [17], [18] /.

Információt kell szolgáltatni a felsőszintű és középszintű /megyei/ vezetés számára a fekvőbetegellátó intézetek működéséről, a hospitalizált betegek összetételéről és ellátottságáról. Ezek az informá-

ciók szükségesek a kórházi ellátottság felméréséhez, az irányításhoz és a tervezéshez. Az információszolgáltatás a következő tényezőkre kell hogy kiterjedjen. A betegek kor, nem, foglalkozás, lakóhely, stb. szerinti megoszlása, az egyes területek ellátottsága, a különböző betegségek elterjedése, az ápolási idő megoszlása az egyes területek, szakmák, diagnózisok között, stb. Az ápolási idők jól jellemzik az egyes területek ellátottságát és az egyes intézmények működésének hatékonyságát. Ugyancsak fontos vizsgálni az egy főre jutó ellátottság mértékét /10 ezer lakosra jutó ápolási eset és nap/.

Fontos szerepet játszanak a korábbi hazai és a nemzetközi adatokkal /WHO statisztikákkal/ való összehasonlítások is. Ezekből következtetni lehet az ellátottság színvonalára és a fejlődés trendjére. Ezért fontos a vizsgálatok folyamatos ismételése. Az összehasonlíthatóság érdekében az információkat megfelelő formában kell szolgáltatni.

Az orvosi gyakorlatban a tisztán gyógyító munka mellett egyre jelentősebb szerepet játszik a megelőzés is. Ehhez ismerni kell a betegségek elterjedését és a gyógyítási /megelőzési/ eszközöket lehetőségeket.

A kórházi morbiditási vizsgálaton belül néhány tényező külön kiemelhető. Az ápolási idők fontosságáról már szóltunk. Külön vizsgálat tárgyát képezhe-

tik a folyamatosan /pl. egy évnél tovább/ ápoltak, a külföldi állampolgárok, a többször ápolott személyek, egy /vagy több/ ápoláson belül jelentkező különböző diagnózisok kapcsolata, a halálozások, a műtéti tevékenységek vizsgálata, az egyes területek /megyék/ közti betegáramlás követése.

1.3.2 Módszertani kérdések. Módszertani problémákról a következőket mondhatjuk el. Ezek egyrészt orvosi, szervezési jellegűek /ld. pl. [10], [17], [18] /, másrészt számítástechnikai és matematikai problémák /[9], [10], [11], [15], [26] /.

A szervezési kérdések közül itt a következőket emeljük ki /egyéb problémákkal bőségesen foglalkozik pl. [8], [18], [19], [25] /. Meg kell fogalmazni a vizsgálat céljait. Ebben sok segítséget nyújtanak az eddigi felmérések eredményei, de a kérdést újra és újra fel kell vetni. Mivel számológépes feldolgozásról van szó, ezért az információigényeket pontosan meg kell határozni. Ugyancsak pontos fogalmazásra van szükség az egyes kódrendszerek, fogalomkörök leírásában is. Itt külön feladatot jelent a más rendszerekkel való kapcsolódás biztosítása. A kórházi morbiditási rendszeren belül nagy figyelmet kell fordítani az egyes diagnózisbesorolások egyértelműségére /fő kórisme, kísérő betegségek, a halál okának különböző megközele-

tései/, az ápolási eset - többszörös ápolás - áthelyezés - új felvétel kérdéskörre, a kórházak és szakmák egy tágabb fogalomkörbe /egészségügyi intézmények és munkahelyek/ való illesztésére.

Számítástechnikai, matematikai /statisztikai/ szempontból a következő kérdések kerülnek előtérbe. Az első kérdés az, hogy milyen módon közelítsük meg a feladatot, milyen számítástechnikai modellt alkalmazunk. Nem csak egészségügyi, de más rendszerek számítógépes vizsgálatakor is több különböző feladattípus van. Az egyik csoportba olyan egyedi számítástechnikai feladatok tartoznak, amelyek célja valamilyen egyszeri tudományos vagy tervezési feladat számítógéppel segített megoldása. Orvosbiológiai, szervezési alkalmazások közül példa erre mondjuk egy idegéletteni jelenség modellezése, egy adott betegség epidemiológiai vizsgálata vagy egy létesítendő kórház optimális méretének és telepítési helyének megválasztása / a SZTAKI-ban végzett ilyen jellegű munkák közül az idősorelemzés alkalmazását EEG jelek alapperiódusának becslésére, szívinfarktus gyakoriság heti és napi periódusának megállapítására, epidemiológiai vizsgálatokat légzőszervi megbetegedésekkel kapcsolatban, gyógyszerhatástani kutatásokat, genetikai vizsgálatokat, stb. említhetjük - ilyen irányu eredményeinket például a hagyományos szegedi kollokviumon

is számos előadásban publikáltuk, ld. pl. [7], [16] /. Ezeket az egyedi vizsgálatokat persze el lehet végezni általánosan használható software eszközök /pl. programcsomagok/ segítségével is /idősoranalízisnél ld. pl. [16], [31], [37] /. Egy másik feladatkör nagy adatbázisok létrehozása, karbantartása és egyedi adatok kisebb vagy nagyobb tömegű lekérdezése /intézetünk ezen a téren is bőséges tapasztalatokkal rendelkezik, pl. a Dunai Vasmű számára készített rendszerek, vagy egészségügyi vonalon az Infarctus Regiszter - ld. [28] - emlithető/. A kórházi morbiditási vizsgálatokban elsősorban nem speciális, egyedi vizsgálatokat kell végezni, nem egyedi adatok nyilvántartását kell biztosítani, hanem nagy tömegű adat egyidejű, jól áttekinthető megjelenítését. A hospitalizált betegekről tehát statisztikákat kell szolgáltatni - ezen az alapon kell a számítógépes rendszert is kidolgozni.

1.3.3 A számítógépes rendszer. A kórházi morbiditási vizsgálat előzőkben összefoglalt céljait és az 1972-73. évi feldolgozás tapasztalatait figyelembe véve, a következő rendszerkép alakul ki. Megjegyezhető, hogy jelenleg már az 1974-77. évi feldolgozások tapasztalatai is rendelkezésünkre állnak - ezeket a munkákat az ESZTIK végezte el, már a SIS77 segítségével.

A kórházi morbiditási vizsgálat nagy mérete és

bonyolultsága miatt az információs igényeket nem lehet néhány előre rögzített statisztikai táblázattal akár csak rövid távon is kielégíteni. Ezt a kijelentést számos tény igazolja. Már az 1972-73. évi vizsgálatnál is /ahol eredetileg rögzített táblatervek szerepeltek/ sok módosításra volt szükség. Ezt a problémát már akkor is egy rugalmas lekérdező rendszer segítségével oldottuk meg /ld. [11]/. A vizsgálat kiszélesítésével ez a tendencia csak fokozódhat. A változékonyságnak vannak azonban sokkal kényszerítőbb okai is. Megváltozhat például a vizsgálat alapját képező adatrendszer. A kórházi morbiditási vizsgálaton belül most van éppen folyamatban az új adatszolgáltató rendszer kialakítása. Ennek keretén belül lényegesen módosul pl. a jelenlegi kórházkód, a szakmák és a beteg foglalkozásának kódolása, és más lesz a betegazonosító kód is. Külső tényezőket is figyelembe kell venni. A közeljövőben kerül bevezetésre az ÁNH azonosító - a betegazonosításra ezt kell majd használni. A betegségek nemzetközi osztályozását /BNO/ is rendszeresen átdolgozzák. Az így megváltozott adatok esetenként teljesen új ellenőrzési eljárásokat igényelnek, és megváltozhat az adatok között fennálló logikai kapcsolatok rendje is. Az adatváltozás eredményeként az információs igények is változnak. Ugyancsak változás forrásai lehetnek a rendszer működtetése

/az információk értékelése/ folyamán nyert tapasztalatok. Számítani kell a feldolgozást igénylő szervezet /egészségügyi vezetés, kutatóhelyek/ módosulásából eredő információigény változásokra is.

Egy másik fontos szempont az, hogy az információigények kialakítását nem célszerű egyetlen lépésben elvégezni. Ehelyett egy szekvenciális feldolgozásmód javasolható /ld. [17]; a "szekvenciális" szót a statisztikai döntéseméletből kölcsönöztük, nem tévesztendő össze a szekvenciális file kezeléssel/. A betegforgalom legfontosabb mutatóit tartalmazó táblázatokon, és a hagyományosan bevált statisztikák szolgáltatásán túl először csak kisebb terjedelmű, durvább bontásban készülő táblázatokra van szükség - esetleg nem is a teljes minta alapján. Ezekből az alapstatisztikákból kiindulva alakítható ki a részletesebb, mélyrehatóbb információigény megfogalmazása. Ilyen szekvenciális feldolgozásmód természetesen nem képzelhető el egy előre rögzített rendszerben.

Az eddig felsorolt szempontokon túl még egy lehetőségre is fel kell hívni a figyelmet. A SIS77 rendszer /és az 1977 augusztusától működő részrendszere is/ igen részletes információszolgáltatást tesz lehetővé, pl. akár egyes kórházak vagy kutató orvosok számára is.

Az eddigieket összefoglalva tehát mindenképpen egy a változó igényekhez rugalmasan alkalmazkodó, ál-

talános statisztikai adatfeldolgozó rendszerre van szükség. A kórházi morbiditási információs rendszer nagy mérete miatt a feldolgozás optimalizálását is meg kell oldani.

A kórházi morbiditási rendszerből eredő igények felsorolása után most összefoglaljuk azt, hogy milyen feladattípusokat kell a SIS77 rendszernek megoldania. Tetszőleges fix adatrekord esetén biztosítani kell az adatok részletes számszerű és logikai ellenőrzését és új adatok képzését /pl. a diagnózisok nemmel, korral való összeférhetetlenségének vizsgálátát, a négyjegyű betegségkódokból különböző összevont jegyzékek - A, B, C, D jegyzék - főcsoportkódok képzése/. A rendelkezésre álló adatokból tetszőleges kombinációban statisztikai táblázatokat kell készíteni. A táblázatokon belül meg kell oldani tetszőleges összevonások /pl. korcsoportok, ápolási időcsoportok/ és részösszegek /pl. összes fizikai dolgozók, összes szellemi foglalkozásuak, stb./ létrehozását, az adatok százalékos megoszlásának kiírását. Az ápolási idő kiemelt szerepe miatt a legtöbb esetben az ápolási napok számát, százalékos megoszlását és az átlagos ápolási napot is közölni kell. /Az ápolási idő a kórházi morbiditási vizsgálat speciális fogalma. A SIS77 rendszer természetesen képes tetszőleges additív statisztika kiszámítására./ Ugyancsak

fontos az egy főre jutó ellátottság /10 ezer lakosra jutó ápolási eset és nap/ kimutatása is. A statisztikákat szemléletessé teszi az adatok grafikus megjelenítése. A rendszer meg kell hogy oldja a többszörös ápolások illetve a különböző diagnózisok kapcsolatának vizsgálatakor felmerülő speciális problémákat is. Ezekben az esetekben azonban nem lehet közvetlenül felhasználni az ápolási esetekre vonatkozó alaprekordokat, ezért bizonyos file-módosításokra van szükség.

Külön kell szólni a rendszerhez kapcsolódó matematikai /statisztikai/ problémákról /ld. [15], [23], [26]/. Ezek egyrészt a személyazonosító kód kialakításával kapcsolatosak, másrészt az optimális nagyságú minta meghatározására illetve a statisztikák megbízhatóságára vonatkoznak. Matematikai statisztikai eszközök használhatók a számítógépes rendszer tervezésénél és alkalmazásánál is /ld. pl. [11], [23], [29]/. A számítástechnikai eszközök alkalmazásáról a SIS77 részletesebb bemutatásakor, a következő fejezetben szólnunk.

1.4 A SIS77 kialakulásának folyamata

A SIS77 legfőbb alkalmazási háttere a kórházi morbiditási vizsgálat. Ezért döntő jelentőségű volt a rendszer szempontjából az Egészségügyi Minisztérium 1973. évi döntése, melyben a kórházi morbiditási vizsgálatok folyamatossá tételét határozták el. /A jelenlegi felméréseknél a mintavételi arány 10 %./ A folyamatos vizsgálatához azonban egy olyan jól működő feldolgozó rendszerre van szükség, amely kielégíti az előző pontokban felsorolt igényeket. Az 1972-73. évi felmérés egy módszertani kísérlet volt, így a számítógépes feldolgozást sem lehetett egy egységes koncepció alapján felépített rendszerrel elvégezni. Szükség volt tehát a kórházi morbiditási vizsgálat számítógépes háttérének megteremtésére.

Az 1972-73. feldolgozás tapasztalatai alapján, részben az ott jelentkező igények kielégítésére már elkészült egy általánosabb alapokon nyugvó részrendszer /ld. [11], [15], [26]/. Ilyen irányu első eredményeinket bemutattuk a WHO Budapesten rendezett 1974. évi Statisztikai Vándorszemináriumán is /[21]/. Ebből a részrendszerből kiindulva a SZTAKI 1975. februárjában részletes javaslatot tett egy általános statisztikai adatfeldolgozó rendszer kialakítására.

Elképzeléseinket még ugyanezen év májusában bemutattuk az Egészségügyi Minisztérium és az ESZTIK munkatársainak. Ezen az összejövetelen - egy általános tájékoztatáson túl - az Akadémia CDC 3300-as gépén működő rendszer táválomáson keresztül történő felhasználását láthatták a résztvevők. A helyben feltett kérdésekre /ezek különböző statisztikai táblázatok igénylését jelentették/ a rendszer még a bemutató ideje alatt választ adott.

Az ÁSZSZ Honeywell 66/20-as és 66/60-as gépének beállításával az egészségügyi vezetés rendelkezésére állt már a feldolgozásokhoz szükséges gépkapacitás is. Így került sor 1976 novemberében az ESZTIK és a SZTAKI közötti megállapodásra. A rendszerkialakítás alapja /ugyanugy mint az 1972-73. évi feldolgozásnál/ a két fél rendszeres konzultációja volt. Ez a kapcsolat és a régebbi tapasztalatok biztosították azt, hogy a SIS77 maximális mértékben igazodjék a kórházi morbiditási vizsgálat igényeihez. Az előző pontokban vázolt szempontok miatt azonban mindenképpen egy általános feltételek között működő rendszerre volt szükség.

Elképzeléseink helyességét és az együttműködés hatékonyságát bizonyítja az, hogy az előirányzott idő előtt már 1977 augusztusában megindulhatott a

kórházi morbiditási adatok tényleges feldolgozása. Két hónappal később már elkészültek az első feldolgozandó év /1974/ táblázatai is /ezek a táblázatok a későbbiek folyamán természetesen még kiegészültek, módosultak/. A feldolgozásokat /a SZTAKI támogatásával/ az ESZTIK végezte.

A SIS77 rendszer jelenleg már lehetővé teszi azt, hogy előre rögzített táblatervek esetén egy-egy újabb évi anyag feldolgozása - leszámítva a hosszadalmas adatfelvételt és adatrögzítést - akár 50-100 nagyméretű táblázat készítésekor is néhány nap alatt véghezvihető. Ezt tényleges futtatási tapasztalatok /az 1974-76 évi feldolgozások/ bizonyítják. Ezekhez a feldolgozásokhoz alacsonyabb képzettségű kezelőszemélyzet is elegendő. Új feladatok megfogalmazása esetén - már csak az orvosszakmai kérdések áttekintésének, megértésének érdekében is - természetesen megfelelő képzettségű, és a rendszer működését, lehetőségeit, korlátait jól ismerő szakemberre van szükség. Ilyenkor ugyanis az egészségügyi szervezés és az orvostudomány területén megfogalmazott újszerű problémáknak számítástechnikailag kezelhető formában történő megfogalmazására van szükség - még egy magas szinten automatizált rendszer esetén is.

1.5 A SIS77 alkalmazásának továbbfejlesztési lehetőségei

A SIS77 általános statisztikai adatfeldolgozó rendszer, így a kórházi morbiditási felméréseken túl más adatrendszerek feldolgozását is lehetővé teszi.

A rendszer kényelmes feldolgozást és gyors átfutást biztosít, így különösebb gépidő és munkaerő ráfordítás nélkül is könnyen kiterjeszthető a statisztikai adatszolgáltatás köre. A felső és középszintű vezetés informálása mellett adatokat kaphatnak akár az egyes kórházak is - tehermentesítve a helyi adminisztrációt. A szervezési, irányítási információigényeken túl kielégíthetők az egyes gyakorló vagy kutató orvosok igényei is - pl. lehetőség van akár egyetlen betegség tanulmányozására is. Ezeket a lehetőségeket a gyorsaságon kívül a rendszer nagyfokú kombinációs képessége és magasszintű automatizáltsága biztosítja. Természetesen egy nagyon szerteágazó adatszolgáltatásnál /egyes kórházak, orvosok számára/ csak nagyon tömör információkat /táblázatokat/ célszerű igényelni - pl. egy-egy kórház számára legfeljebb 5-10 ezer nyomtatott sort.

A rendszer gazdaságos működése lehetővé teszi nagy minták /pl. egy teljeskörű kórházi morbiditási

vizsgálat/ feldolgozását is.

Mivel a rendszer nyílt, nem tartalmaz túl sok speciális megkötést, ezért lehetőség van más rendszerekhez való kapcsolódásra.

A megjelenített táblázatok fejlécei - éppen a rugalmas változtathatóság érdekében - tetszőlegesen cserélhetők, így a rendszer különböző jellegű feldolgozásoknál, pl. nemzetközi kapcsolatokban is alkalmazható.

A feldolgozások menetének gyorsítására javasolható, hogy az adatfelvétel és a számítógépes ellenőrzés, javítás folyamatosan - mondjuk havi részletekben történjék. Így elérhető lenne, hogy a kórházi morbiditási vizsgálatokban egy lezárt év vége után már egy hónapon belül rendelkezésre álljanak a szükséges táblázatok. Ez a feldolgozásmód egy folyamatos visszajelzést is lehetővé tesz az adatszolgáltatók felé.

Tanulmányunk egy későbbi fejezetében a SIS77 számítástechnikai vonatkozású továbbfejlesztési lehetőségeivel is foglalkozunk.

2. A SIS77 szerkezete, működési elvei és felhasználási lehetőségei

Ebben a fejezetben megvizsgáljuk, hogy milyen szempontok alapján lehet felépíteni egy statisztikai adatfeldolgozó rendszert. Lényeges figyelembevenni azt, hogy nagyméretű adatrendszerekkel foglalkozunk. Kisebb adattömeg feldolgozására ugyanis rendelkezésre állnak jól kiépített program illetve szubrutin rendszerek. Ezek közül említhetjük a CDC 3300-as gép könyvtári programjairól az MTA SZTAKI kiadásában megjelenő köteteket, a W.J. Dixon szerkesztésében az University of California kiadásában megjelent BMD /Biomedical Computer Programs/ köteteket vagy az IBM 360-as sorozatra kidolgozott szubrutin gyűjteményt /ld.pl. IBM Application Program, System/360, Scientific Subroutine Package, Programmer's Manual/. Ezekben a programgyűjteményekben azonban olyan különálló programok vannak, amelyek egyrészt nagytömegű adat feldolgozásánál nem elég hatékonyak, másrészt nem biztosítanak egy egységes adatfeldolgozási folyamatot, hanem csak bizonyos tipikus táblázási és statisztikai feladatokat oldanak meg. Ehhez járulnak még a - rendszer

alapját képező - kórházi morbiditási vizsgálat sajátos igényeiből adódó speciális file-szerkesztési feladatok és táblázatformátumok.,

A most következő fejezetben általános számítástechnikai, szervezési, matematikai, statisztikai kérdésekkel foglalkozunk. Alkalmazási vonatkozású kérdésekben azonban mindig a kórházi morbiditási vizsgálatra hivatkozunk, és azt mutatjuk meg, hogy miként építhető fel egy ennek az adatfeldolgozási rendszernek megfelelő számítógépes rendszer. Mivel a SIS77 jelenlegi megvalósítása az ÁSZSZ HWB 66/60-as /és 66/20-as/ gépén történt, ezért a Honeywell rendszer által adott lehetőségeket kihasználjuk.

A jelen tanulmány elsősorban általános vonatkozású kérdésekkel foglalkozik. Az egyes részrendszerek, számológépprogramok részletes leírására egy külön tanulmányban kerül sor.

2.1 A statisztikai adatfeldolgozás egy modellje

2.1.1 Adatbázisok - statisztikai adatfeldolgozás.

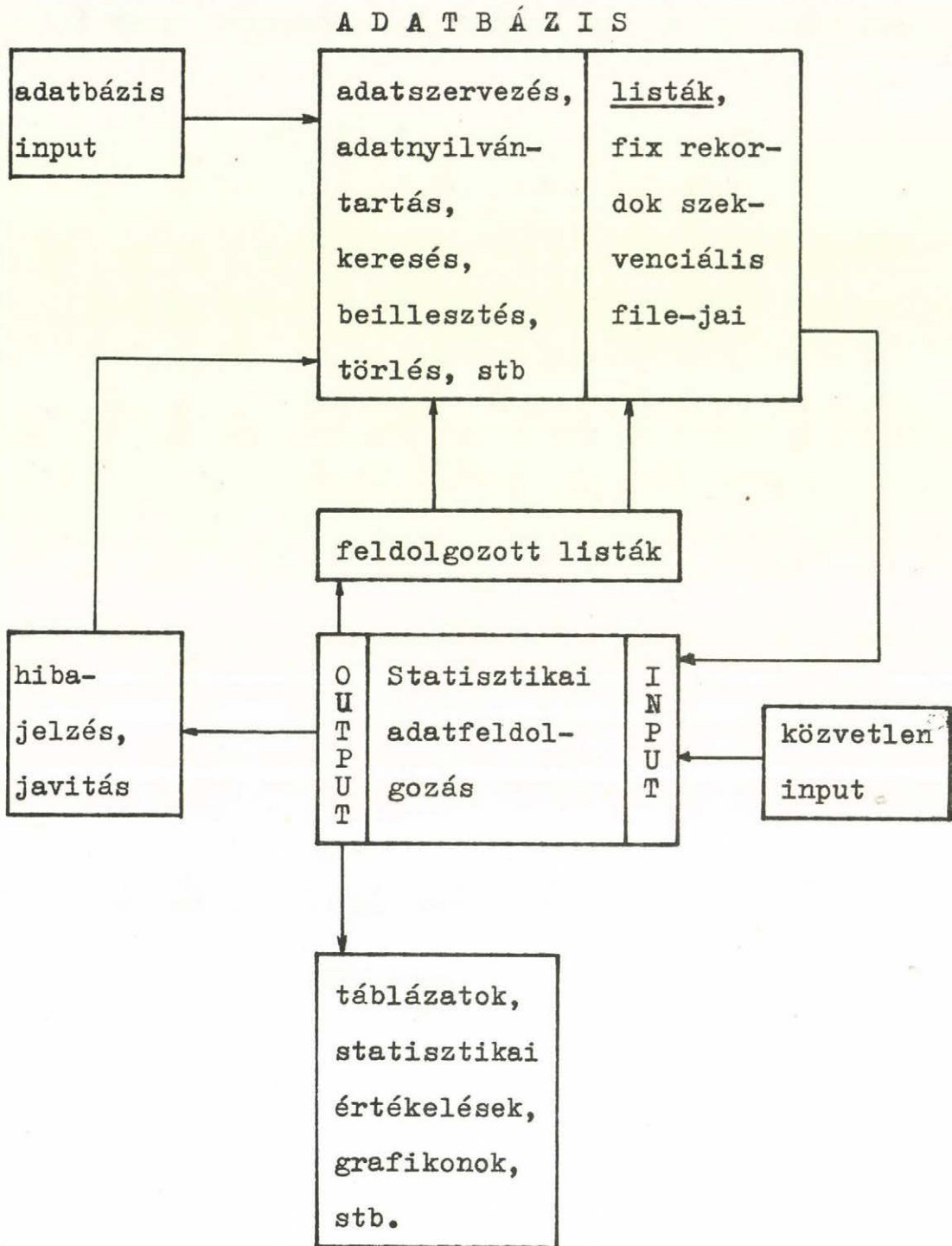
A statisztikai adatfeldolgozásnak, az adatbázisok fogalomköréhez kapcsolódva sajátos fogalmai és problémái vannak. Ezeknek a fogalmaknak és kérdéseknek a

vizsgálatával el lehet jutni azokhoz a módszerekhez, amelyek lehetővé teszik általánosan alkalmazható statisztikai adatfeldolgozó rendszerek kialakítását. A SIS77 felépítésénél különös figyelmet fordítottunk még a nagyméretű adatrendszerek speciális problémáira és a kórházi morbiditási rendszer igényeire is.

Adatbázisok létrehozásakor a különböző szintű problémák egész sora merül fel. Ki kell alakítani azt a fogalomkört amelyben az adatfeldolgozást folytatni kívánjuk. A létrehozott fogalomkörhöz egy rendszermodell kell konstruálni, és a rendszermodell számológépes megvalósítását is el kell végezni - miközben sok olyan technikai részletkérdést is meg kell oldani, amelyek komoly számítástechnikai és matematikai felkészültséget igényelnek /pl. rendezési, keresési, file-szervezési, stb technikák megalkotása/. Ilyen általános kérdésekkel számtalan publikációban találkozhatunk /ld. pl. [2], [4], [5], [12], [20], [24] /.

Adatbázisok és statisztikai adatfeldolgozások egy lehetséges kapcsolatrendszerét szemlélteti az 1. ábra. /Ezzel a kérdéskörrel foglalkozik még [32], [33] /.

Nagyméretű adatfeldolgozási feladatoknál első-sorban a legegyszerűbb file-típusok - listák /fix rekordok szekvenciális file-jai/ - használata gazdaságos, másrészt a statisztikai adatfeldolgozásokat sokszor egy adatbázis kialakításától függetlenül,



1. ábra

egy adott állapotra vonatkozóan kell elvégezni. Nagytömegű adat statisztikai vizsgálatát célszerű tehát valamilyen módon leválasztani egy általános adatbázisról. Ezt mutatja az 1. ábrán a "listáknak" az adatbázis egészétől való elkülönítése.

Az adatbázisok szervezését és lekérdezését megvalósító nyelvek, rendszerek /pl. a Honeywell rendszerben az IDS és az MDQS/ elsősorban viszonylag nem túl nagy, de bonyolult szerkezetű adatbázisok kezelését és egyedi adatok elérését /ld. pl. helyfoglalás repülőjáratokra, bankszámla lekérdezés, stb/ teszik lehetővé. Ezért ezeket a rendszereket nagyméretű, de viszonylag egyszerű adatszerkezettel dolgozó statisztikai adatfeldolgozásokban nem célszerű alkalmazni, már csak azért sem, mert a statisztikai adatfeldolgozások nem egyedi adatokkal, hanem tipusokkal foglalkoznak, és sajátos output formátumokat /táblázatok, grafikonok/ igényelnek.

A statisztikai adatfeldolgozásnak tehát egy viszonylag önálló fogalomköre és eszköztára van. Ezért választottuk el élesen az 1. ábrán bemutatott modellben az általános adatbázistól. A feldolgozó rendszer inputja természetesen származhat egy adatbázisból is, és a statisztikai feldolgozások közbeeső fázisaiban részeredmények /szintén egyszerű listák/ ugyancsak

tárolhatók egy általános adatbázisban. Tipikusan adatbázis szerkesztési feladatok a hibás rekordok javításával, új rekordoknak az adatfile-ba történő illesztésével, különböző file-ok egyesítésével kapcsolatos munkák is. A cél azonban mindig az, hogy az adatbázis a statisztikai feldolgozás számára mindig egy egyszerű adatszerkezetet adjon át.

2.1.2 Egy modellvázlat. Az 1. ábrán bemutatott vázlat tárgyalása után most szűkítsük le vizsgálatainkat a statisztikai adatfeldolgozás területére. Milyen fogalmak, feladatkörök kell hogy szerepeljenek egy általános statisztikai adatfeldolgozó rendszerben?

2.1.2.1 Az egyik első lépés a felhasznált adatok ellenőrzése, kiegészítése. Az általánosan is használható könyvtári programok között ezen a területen csak a legegyszerűbb vizsgálati lehetőségek állnak rendelkezésre. Ezért a legtöbb adatfeldolgozási feladatnál a felhasználó arra kényszerül, hogy önállóan készítsen adatellenőrző programot. Ez a megoldásmód viszont rendkívül munkaigényes, lelassítja a feldolgozást és sok hibalehetőséget rejt magában. Ez a kérdés különösen fontos, hiszen csak megfelelően előkészített és ellenőrzött adatokkal lehet csak a feldolgozás további lépéseit elvégezni. Mindezeket figyelembe véve elmond-

ható, hogy bármely adatfeldolgozó rendszernél feltétlenül szükséges egy előkészítő részrendszer, amely viszonylag bonyolult adatellenőrzéseket, adatértékek logikai kapcsolatainak ellenőrzését is lehetővé teszi. Fontos szempont az, hogy bár az ellenőrzési eljárások leírását mindenképpen a felhasználónak kell megadnia, - hiszen saját igényeit csak ő ismeri - de a rendszernek egy olyan részletes és jól áttekinthető diagnózist /hibajelzést, értékelést/ kell készítenie a felhasználó leírásáról /a feladat kódolásáról/, hogy abban a hibalehetőségek minimálisra csökkenjenek. Ugyancsak lényeges az, hogy a felhasználó a szokásos programozási nyelvek által adott lehetőségeknél sokkal kényelmesebb módon írhasse le az ellenőrzési feltételeit - hiszen egy speciális feladatkört megoldó rendszerről van szó. Ez a megállapítás a következőkben leírásra kerülő más részrendszerekre is igaz, tehát az adatfeldolgozás folyamatának leírására egy ahhoz jól alkalmazkodó célnyelvet /paraméterezési rendszert, utasításrendszert/ kell kialakítani. Ez az elv természetesen más számítástechnikai feladatnál is alkalmazható.

A statisztikai adatfeldolgozás előkészítő fázisában az ellenőrzés mellett szükség lehet új adatok képzésére is. Az adatfelvételnél ugyanis felesleges olyan adatokat rögzíteni, amelyek más, már rögzített adatok függvényei /pl. a diagnózisok A, B, C... jegyzék szerinti kódszámát egyértelműen meghatározza a négyjegyű betegségkód/. Vegyük észre, hogy az új adatok létrehozása a hibavizsgálat egy általánosítása, hiszen a hibakeresés is egy speciális

függvény, a "hibás", "hibátlan" értékkészletű függvény kiszámítását jelenti /a hibakeresést felfoghatjuk mint egy speciális átkódolási feladatot: pl. a jó adatoknak az 1, a rossz adatoknak a 0 értéket adjuk/. Ezért mindkét esetben hasonló módszerek alkalmazhatók /ld. [36]/. Az ellenőrzést és az új adatok létrehozását meghatározó logikai strukturák leírásában illetve programozásában fontos szerepet játszanak a különböző keresési eljárások /erről a kérdéskörrel ld. pl. [20], [36]/.

2.1.2.2 Adatfeldolgozásnál, és így a statisztikai vizsgálatoknál is, sokszor a teljes adatrendszer egy részére van csak szükségünk, ezért egy válogatási eljárást kell végrehajtani. Ez egyaránt lehet az adatrendszer egyes rekordjainak vagy az egyes rekordokon belül kijelölt adatmezőknek a válogatása. Meg kell jegyezni, hogy a válogatás is egy speciális átkódolást /függvény képzést/ jelent, hiszen a "szükséges" és "nem szükséges" "értékeket" kell meghatározni. Bonyolultabb esetekben mind az ellenőrzés, mind az adatrendszer felbontás /válogatás/ sok érdekes számítástechnikai /statisztikai, logikai, gráfelméleti/ problémát vet fel /ld. [23], [29], [30], [36]/. Az adatrendszer felbontása elsősorban a feldolgozás hatékonyságának javítása érdekében történik. Erről a kérdéskörrel a következő /2.2/ pontban részletesebben beszélünk.

2.1.2.3 A rendszermodellben fontos szerepet játszanak a felhasznált software-eszközök, programozási nyelvek és adattárolási módok. Ezekkel a kérdésekkel ugyancsak a 2.2 pontban foglalkozunk.

2.1.2.4 Mint ahogy már az előzőkben is hangsúlyoztuk, a statisztikai adatfeldolgozásokban nem egyedi adatok, hanem az egyedek típusainak adatai szerepelnek. Ezért a feldolgozás során nem az egyedek adatait kell tárolni illetve vizsgálni, hanem a kívánt típusokat. A típusok elsődleges statisztikai jellemzője előfordulási gyakoriságuk. Ehhez társulhat valamilyen más statisztikai adat - a kórházi morbiditási vizsgálat esetében például az ápolási idők összege, a műtétek számának összege, egy típuson belül a meghaltak száma, stb. Ezeket az értékeket /gyakoriságok, összegek/ táblázatokban tárolhatjuk. Az egyedek adatait tartalmazó file-okról /rekordlistákról/ áttérünk a típusok adatait tartalmazó file-okra /táblázatokra/. Az ilyen transzformáció után nyert file-okat nevezhetjük transzformált file-nak, táblafile-nak, statisztikai file-nak vagy a típusok file-jának. A következőkben a táblafile elnevezést használjuk. A táblafile-okból képezhetjük a megjelenítésre /nyomtatásra, távállomás képernyőjére/ kerülő táblázatokat, és a táblafile-ok értékeivel végezhetünk statisztikai számításokat is.

A táblafájl alkalmazásának elvét - ugyanugy mint az adatrendszer felbontásában rejlő hatékonyságnövelő lehetőségeket - már rendszerünk kiépítésének legelső fázisában is felhasználtuk /ld. [11], [15], [22], [26] /.

2.1.2.5 A tisztán statisztikai adatfeldolgozási feladatokon belül is szükségesek olyan lépések, amelyek kifejezetten adatbázisszerkesztési eljárásokat igényelnek. A 2.1.1 pontban említetteken kívül itt a kórházi morbiditási rendszer két feladattípusára hívjuk fel a figyelmet. Az egyik a kísérő és következményes betegségek, a másik a többször ápolott személyek vizsgálata. Mindkét esetben nem az ápolási eseteket képviselő rekordok file-ját, hanem egy abból képzett file-t használhatunk, amelyben az egyes rekordok a különböző diagnózisokhoz illetve személyekhez tartoznak. Ez a feladatkör szorosan kapcsolódik az ugynevezett relációs adatmodellek kérdésköréhez / [5], [33] /.

A szükséges adatbázisszerkesztési feladatokat beépíthetjük statisztikai rendszerünkbe is /a SIS77-ben ezt a megoldást választottuk/, de egy már meglévő adatbázison keresztül is eljuthatunk a kívánt új file-szerkezetekhez. Ez a kapcsolat az 1. ábrán vázolt formában úgy képzelhető el, hogy a statisztikai rendszer által kezelt listákat /pl. az ápolási esetek rekord-

jait/ veszi át az adatbázis akár tárolás, akár egy átszerkesztés céljából. A statisztikai rendszer az adatbázisból mindig csak fix rekordu szekvenciális fileokat /listákat/ kap.

2.1.2.6 Az adatfeldolgozó rendszereknek a felhasználó oldaláról nézve egyik legfontosabb része az eredmények megjelenítése /kinyomtatása, képernyőre vitele/. A statisztikai feldolgozásokban a felhasználó elsősorban gyakoriságeloszlásokat és a megfelelő százalékos eloszlásokat /relatív gyakoriságokat/ igényel.

Egy másik elég gyakran jelentkező igény az, hogy bizonyos kódösszegek - százalékos eloszlásukkal és átlagos értékükkel együtt - is megjelenjenek. A kórházi morbiditási vizsgálatban szerepelhet pl. az ápolási napok összege, stb /ld. 2.1.2.4/ és az átlagos ápolási nap. Más esetekben lehet összjövedelmet, összteljesítményt, egy főre jutó jövedelmet vagy teljesítményt számítani. Az ilyen additív statisztikák - összegek, szorzatösszegek, négyzetösszegek, stb - további statisztikai számítások, például a szokásos többváltozós analízisek alapját képezhetik.

Az egyéb - programkönyvtárakban is megtalálható - statisztikai feladatokat, például eloszlás-vizsgálatokat is ebben a fázisban lehet elvégezni.

Ha a vizsgálat nem egy teljes populációra, hanem annak egy speciális részére /pl. csak a kórházban

ápolott betegek/ vonatkozik, akkor fontos lehet a kiemelt rész és az egész viszonyának vizsgálata. A kórházi morbiditási vizsgálatban ilyen céllal készülnek kimutatások pl. a 10 ezer lakosra jutó ápolási esetről és napról. A viszonyítás történhet egy a vizsgált populációtól eltérő halmazzal is, pl. az egy kórházi ágyra jutó betegek számának meghatározásakor.

Az adatok megjelenítésének fázisában is fontos a variálhatóság biztosítása. Az a cél, hogy a felhasználó a kapott eredményeket közvetlenül alkalmazhassa - tartalom és forma tekintetében egyaránt. Ezért a statisztikai táblázatoknál gondoskodni kell a változó tartalom mellett a megjelenítési forma változtathatóságáról is, pl. egy adott határon belül tetszőlegesen megválasztható táblázat-dimenziószám, cserélhető fejlécszövegek, részösszegek, összevonások, kihagyások automatikus biztosítása, gyakoriságeloszlás helyett kumulatív értékek kiírása, stb.

Az eredmények gyors áttekintését nagyban segíti az értékek grafikus megjelenítése.

2.1.2.7 Szólni kell még néhány kiegészítő feladatról is. Adatfeldolgozás folyamán a különböző transzformációkat, ellenőrzéseket leíró táblázatokat a felhasználó kell hogy kitöltse. Célszerű ezeket az adatbeviteli munkákat egy kényelmesen és biztonságosan

használható programmal segíteni /ld. [32] /. Ilyen feladatokra jó példa a négyjegyű BNO kód /betegségek nemzetközi osztályozása/ tömörebb jegyzékekre, például a 300 elemű D jegyzékre való leképezése. Ebben az esetben tizezerféle kódot kell egy 300-féle kódot tartalmazó halmazra leképezni. Ezt a leképezést a BNO négyjegyű kódjának és a D jegyzéknek bonyolult kapcsolata miatt csak értéktáblázattal lehet megadni, amit mindenképpen a felhasználó kell hogy kitöltsön.

Nagyobb rendszereknél feltétlenül hasznos egy adminisztráló részrendszer alkalmazása. Egy ilyen részrendszer bármely adatfeldolgozási feladatnál segítheti a felhasználót a létrehozott részeredmények, a közbeeső állapotok, file-ok biztonságos nyilvántartásában és az egyes részrendszerek közti kommunikációban.

Szükség lehet még a feldolgozandó adatok file-jainak tárolása mellett kiegészítő adatok tárolására is, például a táblázatok fejlécszövegeit, az adatértékeket, a viszonyításokhoz felhasznált statisztikai alapadatokat lehet ezeken a segédfile-okon elhelyezni.

2.1.3 A modell megvalósítása. Hogyan foglalhatjuk az előző /2.1.2/ pontban vázolt feladatköröket egy egységes rendszerbe? Milyen szempontokat vegyünk figyelembe

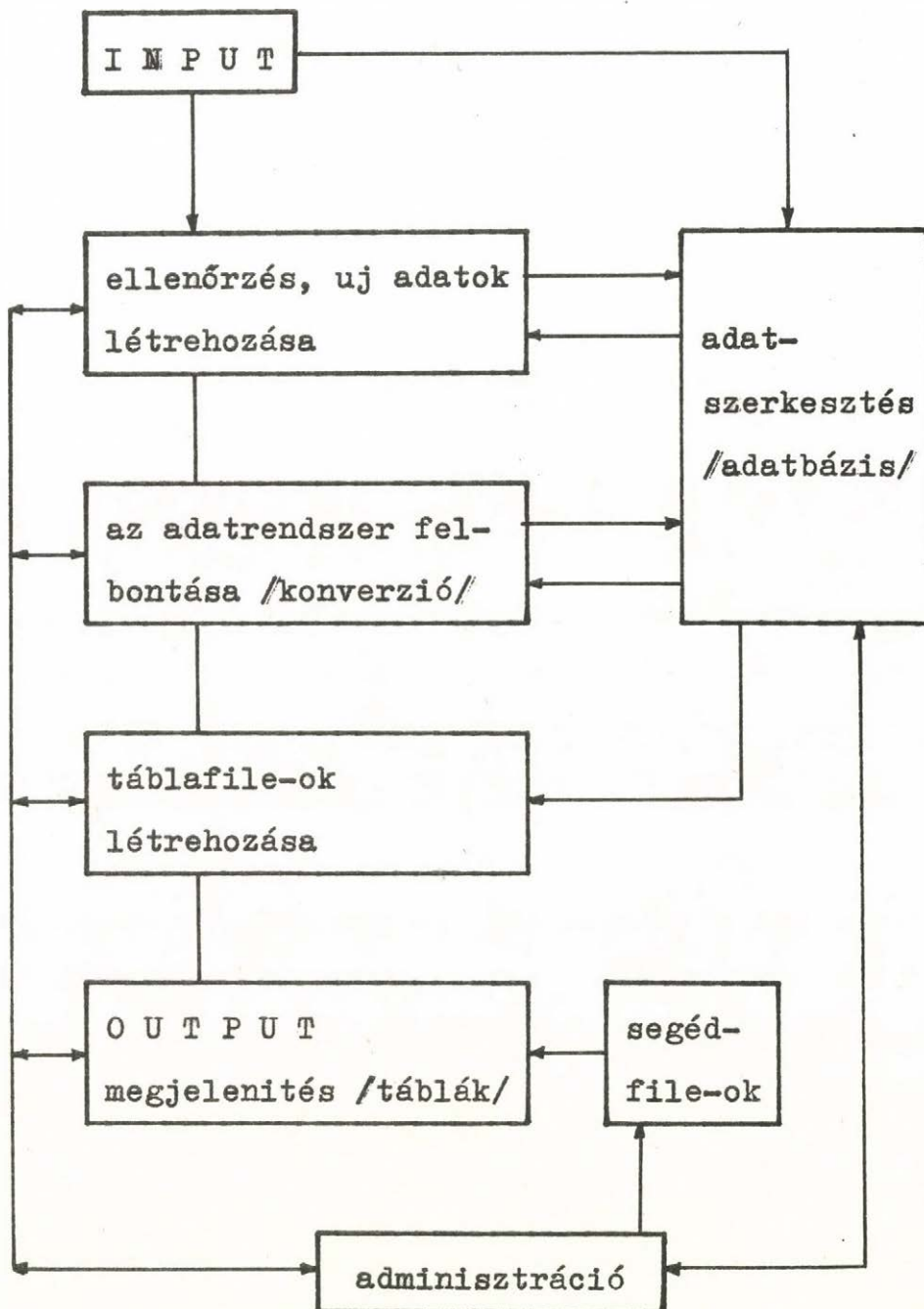
a rendszer számológépes megvalósításakor? Felsorolunk néhány olyan szempontot, amelyek más /adatfeldolgozási/ rendszereknél is figyelembe vehetők.

A rendszer legyen teljes, a kitűzött feladatok megoldásához lehetőleg ne legyen szükséges több, különböző helyről származó független rendszert /programot/ alkalmazni /a kompatibilitási problémák elkerülése érdekében/. Ugyanekkor a rendszer legyen nyílt, más rendszerekhez való kapcsolódás lehetőleg minél kevesebb problémát okozzon. A rendszerben megoldható feladatok legyenek többféleképpen kombinálhatók, az egyes részfeladatok legyenek önállóan is megoldhatók. Mindezek a célok elérhetők egy modulszerkezet alkalmazásával. Az egyes modulok közti kapcsolatot /adatátvitelt/ úgy kell megoldani, hogy az egyes részrendszerek /modulok/ egymással vagy egy új rendszerrel /programmal/ is cserélhetők legyenek.

Ezeket a szempontokat és a 2.1.2 pontban írtakat figyelembevéve a következő rendszervázlatot adhatjuk /ld. 2. ábra/.

A rendszer inputja kerülhet közvetlenül ellenőrzésre, kiegészítésre, vagy egy adatbázison /illetve a statisztikai rendszerbe épített szerkesztő eljárásokon/ keresztül érjük el.

Az ellenőrzés és kiegészítés után az adatok vagy közvetlenül, vagy szintén egy szerkesztési eljáráson



2. ábra

/adatbázison/ keresztül jutnak a következő lépésbe. Az utóbbi megoldást például az adatok javítása és a rendszerbe való visszajuttatása, vagy összetett adatrendszerek /ld. pl. integrált egészségügyi rendszer/ esetén választhatjuk.

Az optimális működés érdekében, a feldolgozás közbeeső lépéseiben szükséges file-szerkesztési eljárásokat /ld. pl. a kísérobetegségek és a többször ápoltak esetét, a rendezéseket, a direkt elérésű file-ok létrehozását, stb/ célszerű már a válogatott és konvertált, tehát már a csökkentett méretű és gyorsan feldolgozható file-okon elvégezni. Ezeket a szerkesztési lépéseket újabb válogatás is követheti /pl. egy személyazonosító segítségével összeállított természetes személyek file-jából kiválogathatjuk a többször ápoltakat/.

A táblafile-okat akár közvetlenül egy válogatás /konvertálás/ eredményeként születő file-ból is előállíthatjuk, de lehetséges, hogy előbb valamilyen szerkesztési eljárásra /pl. rendezésre, a többször ápoltszemélyekhez tartozó ápolási esetek egyesítésére, stb/ van szükség.

A rendszeradminisztráció a feldolgozás valamennyi lépését nyomonkövetheti, és a táblázatok készítéséhez felhasznált segédfile-ok feltöltését is célszerű ugyanezzel a részrendszerrel megoldani.

2.2 Optimalizálási lehetőségek

2.2.1 Az optimalizálás jelentősége. Minden termelési, szervezési, kutatási folyamat kialakításakor - és így az adatfeldolgozásban is - egyik elsődleges cél az adott folyamat optimális lezajlására való törekvés. A kórházi morbiditási adatok feldolgozásának esetében két ok is különösen indokolttá teszi az optimalizálási igényt. Egyrészt rendkívül jelentős kérdés vizsgálatáról van szó, hiszen az egész ország kórházhálózatának működéséről kell információkat gyűjteni, másrészt a feladat jelentőségénél és terjedelménél fogva nagy anyagi kapacitást követel, melynél bizonyos arányu megtakarítás nem lehet elhanyagolható. Ugyancsak az optimalizálás szükségességét igazolja az a tény, hogy nem egy alkalmi feldolgozásról van szó, hanem egy állandóan ismétlődő folyamatról /az évenkénti kórházi morbiditási vizsgálatról/. Ezen kívül, mint ahogy már az előzőkben elmondtuk a SIS77 egy általános statisztikai adatfeldolgozó rendszer, tehát széles körben alkalmazható, és így optimalizálása is széleskörű nyereséget jelent. Fel kell még arra is hívni a figyelmet, hogy a statisztikai adatfeldolgozás általános vonatkozásu optimalizálása más számítógépes rendsze-

rek felépítésének, kezelésének optimalizálási módját is megmutathatja. A kórházi morbiditási témakörhöz kapcsolódó optimalizálási kérdésekkel már több előző dolgozatban, tanulmányban foglalkoztunk /ld. [11], [15], [23], [26], [32], [33] /.

Az optimalizálás a feldolgozás különböző oldalaira vonatkozik.

Elsősorban a felhasznált adatrendszerrel a lehető legtöbb hasznos információt kívánjuk nyerni. Ezért az adatfelvételt /a forrásként szolgáló információs rendszert/ gondosan elő kell készíteni, és információigényeinket jól át kell gondolni.

Egy másik tényező az idő. Az időoptimalizálás több vonatkozásban is érdekes: 1. Az adatfelvétel megkezdésétől az eredmények kézhezvételéig eltelt idő minimalizálása. 2. A feldolgozáshoz szükséges számítógépkapacitás - elsősorban a gépidő - csökkentése. 3. A rendszer kialakításához, a programok futtatásához, az eredmények kiértékeléséhez szükséges munkaerőkapacitás - a felhasznált emberi munkaidő - optimális elosztása illetve minimalizálása.

Ugyancsak figyelembe kell venni azt a tényt, hogy a rendelkezésre álló számítógép kapacitása véges. Így ha feladatunk mérete bizonyos korlátokat túllép, kényszerítve vagyunk hatékonyabb eljárásokat keresni.

Mivel az itt vázolt többdimenziós optimalizálási rendszernek a tényezői egyáltalán nem függetlenek, ezért az optimum megtalálása egy bonyolult matematikai modelt igénylő feladat. Tulajdonképpen ezen a területen - mint általában az adatfeldolgozás legtöbb területén - a problémák egzakt megfogalmazása még hátravan.

Az optimumot az egyes dimenziókban sem könnyű megtalálni. Például egy adatrendszerből nyerhető információmennyiség optimalizálásának definiálása sem kézenfekvő - sokszor bonyolult matematikai statisztikai megfontolásokat is igényel /ld. pl. [23]/. Kérdés lehet például, hogy milyen információkat kívánunk nyerni, milyen gyorsan és mekkora erőráfordítással. Ez az egy kérdés önmagában véve is egy többváltozós döntési feladat.

Ezen a területen bármilyen eredmény csak a különböző érintett szakterületek /matematika, számítástechnika, szervezés, orvostudomány, stb/ kutatóinak szoros együttműködésétől várható.

2.2.2 Néhány optimalizálási feladatkör. Az előző

/2.2.1/ pont alapján több különböző feladatkört választhatunk szét.

2.2.2.1 Elsőként a SIS77 alkalmazásának kérdéseivel

foglalkozunk. Az előzőkben /1.3.3 pont/ már volt szó a szekvenciális feldolgozásmód előnyeiről. Ez az adatok áttekinthetőbb, gyorsabb kiértékelését teszi lehetővé / [17] /.

Ugyancsak a felhasználási mód helyes megválasztásával áll kapcsolatban az esetenként szükséges sze-mélyazonosító /rekordazonosító/ kódok megfelelő kialakítása is. Matematikai statisztikai vizsgálatokat végezve lehet egy valóban megbízható, de lehetőleg nem túl bonyolult /tehát optimális/ azonosító kódot konstruálni / [15], [23], [26] /.

Az adatfeldolgozás egyik leglényegesebb pontját alkotják a mintavételi problémák, amikor nem teljes minta esetén a reprezentativitás biztosítása okoz gondot. Itt meg kell azt határozni, hogy az aktuális igényeknek megfelelően milyen szinten kell biztosítani a reprezentativitást. Meg kell határozni, hogy mekkora a kívánt statisztikai megbízhatóságot adó optimális /viszonylag egyszerű módon kiválasztható és nem túl nagy/ minta. Speciálisan a kórházi morbiditási vizsgálatokkal kapcsolatos mintavételi kérdésekkel foglalkozik a [23] tanulmány.

Egy adatfeldolgozó rendszerben megvalósítható az, hogy az adattartalomtól függően választjuk meg az aktuális működésmódot. Ilyen kérdésekkel még a 3. fejezetben is foglalkozunk. Statisztikai táblázatok ilyen készítésének

módon történő optimalizálása, és hatékony adatkeresési /kódolási, javítási/ technikák kialakítása szerepel [11]-ben, [23]-ban és [36]-ban.

Az optimális felhasználhatóság érdekében az egyes programoknál biztosítani kell a változatos paraméterezési lehetőségeket. Táblázatkészítések esetén például különböző táblaformátumokat, részösszegek képzését, kumulatív eloszlás létrehozását, viszonyszámok kiszámítását, grafikus megjelenítést egy-egy paraméter megadásával lehessen igényelni. A SIS77 ezeknek a követelményeknek megfelel.

Az optimális üzemeltetésnek ugyancsak fontos feltétele a megfelelő szintű biztonság. A SIS77 programjai a felhasználó által adott utasításokat /paramétereket/ részletesen elemzik, és a hibás utasításokat, ellentmondó paramétereket kiszűrik. A felhasználó részletes tájékoztatást kap a programok működéséről, a részeredményekről.

Befejezésként még megemlíthetjük, hogy a SIS77 felhasználhatóságát növeli az a tény is, hogy pillanatnyi alkalmazásán /a kórházi morbiditási vizsgálaton/ túl máshol is használható, és mivel egy általánosan használt magasszintű nyelven /FORTRAN/ készült, ezért nem csak a jelenlegi Honeywell rendszerben működhet, hanem más gépekre is átvihető.

2.2.2.2 Egy számítástechnikai feladat megoldásakor a felhasználás mellett a rendszer előállításához szükséges munkaerőt is számításba kell venni. A SIS77 készítésekor két fő szempontot vettünk figyelembe. Az egyik a már előzőekben is említett modularitás. Ez a megoldásmód a feladatok variálhatóságán túl a rendszer előállítását is egyszerűsíti. Az egyes programozási lépések különválaszthatók, és az egyes rendszermodulok több helyen is felhasználhatók.

Egy másik programozástechnikai probléma a felhasznált software-eszközök helyes megválasztása. Egy optimális megoldás - mind a felhasználás, mind a rendszerkialakítás szempontjából - csak a megfelelő eszközök alkalmazásával érhető el. Már az előzőekben felhívtuk a figyelmet arra, hogy statisztikai adatfeldolgozásnál - így a SIS77 rendszerben is - nem egyedi adatok lekérdezését kell megoldani, tehát nem célszerű adatszerkesztő-lekérdező nyelveket, rendszereket alkalmazni.

Hogy az itt említett és a következőkben még szóba kerülő optimalizálási szempontokat kielégítő rendszert könnyen és jól áttekinthető módon létrehozassuk, felhasznált programozási nyelvként a FORTRAN nyelv kiálkozott.

Az optimális megoldásmód igénye indokolta azt is, hogy már létező, de nagyméretű feladatoknál rossz ha-

tásfokkal működő statisztikai programokat nem használ-
tunkfel /ld. ennek a 2. fejezetnek a bevezetését/.

Egyébként is a szükséges output formátumok biztosítása és a speciálisan kórházi morbiditási vizsgálatokban szükséges járulékos adatok megjelenítése ezeknél a standard statisztikai programoknál külön gondot jelentene.

2.2.2.3 Rátérve a szorosan a rendszer működésére vonatkozó optimalizálási kérdésekre, először a programfutásidők minimalizálásával foglalkozunk. A SIS77 rendszerben egy olyan programozástechnikai eszközt dolgoztunk ki, amely lehetővé teszi a futásidők lényeges mértékű csökkentését /ld. [34], [35] /. A nagy kapacitást lekötő feladatok programjait esetenként egy paraméterekkel vezérelhető előkészítő /szerkesztő, generáló/ program állítja elő. Módszerünk tulajdonképpen a számológépfelhasználás hatékonyságnövelésének egyéb problémaköreihez csatlakozik - amelyek közül például a multiprogramozású rendszerek optimális működtetését /([1],[3],[6] /, fordítóprogramok optimalizációs technikáit /([13],[27] / említhetjük.

Ugyancsak jelentős gépidőmegtakarítást eredményez az előzőekben már vázolt modulszerkezet alkalmazása. Ez ugyanis a feldolgozás átgondolt megtervezése esetén lehetővé teszi, hogy bizonyos előkészítő lépé-

seket ne kelljen ismételt elvégezni /ld. 2. ábra/. A SIS77 alkalmazásakor az ellenőrzéseket, az új adatok létrehozását egy feldolgozáson belül általában csak egyszer kell elvégezni; természetesen lehetőség van az input file újabb, kiegészítő ellenőrzésére és módosítására is. Az adatrendszer felbontása és konverziója optimális tárolást és gyors feldolgozást tesz lehetővé /kikerülve a felesleges adatmozgatásokat és az ismételt konverziót/. A táblafileok alkalmazása lehetővé teszi, hogy a statisztikai táblázatokat rendkívül gyorsan /néhány másodperces futásidővel/ megkaphassuk - egy gyakorlatilag tetszőlegesen nagy minta vizsgálata esetén is. Ezekkel, a rendszer felépítésével kapcsolatos kérdésekkel a 2.3 pontban részletesebben foglalkozunk.

Optimális működés megvalósítására törekedtünk az adatellenőrzési, kódolási és válogatási részrendszerek létrehozásakor is. Ezek a részrendszerek tulajdonképpen bonyolult döntési folyamatokat és az adatfeldolgozásban központi szerepet játszó keresési eljárásokat realizálnak, és néhány sztochasztikus optimalizálási problémát is felvetnek /ezekről a kérdésekről ld. pl. [20], [29], [30], [32], [33], [36]/. Ezeknél a feladatoknál a felhasználó szempontjait figyelembevéve kényelmes adatbeviteli eljárásokat biztosítottunk, és

különböző felhasználási opciók megadásával egy optimális üzemmód kiválasztását is lehetővé tettük /ld. [32], [33] /.

2.2.2.4 Ugyancsak a rendszerműködés optimalizálása érdekében alakítottunk ki egy tömör de gyors feldolgozást biztosító tárolásmódot /ld. [32], [33] /. Az adatokat binárisan tároljuk, de nem szavanként, hanem úgy, hogy mindig csak a szükséges számú bitet foglaljuk le. Ez a tárolásmód tömörebb mint például a COBOL karakteres tárolása.

A központi memória tárolókapacitásának optimális /dinamikus/ kihasználását teszi lehetővé az előző /2.2.2.3/ pontban említett programozástechnikai fogás is /a szerkesztő programok alkalmazása/.

A szerkesztő /generáló/ programok alkalmazásával általában egyszerű szerkezetű, jól áttekinthető programokat hozhatunk létre /ld. [34], [35] /, elősegítve ezzel a rendszerelőállítás munkáit /ld. 2.2.2.2 pont/.

Adattárolási kérdésekkel kapcsolatban, speciálisan rendezési eljárások optimalizálási lehetőségeivel foglalkozik [15] és [26].

A rendszer /SIS77/ működésével bővebben foglalkozik a most következő 2.3 pont.

2.3 A SIS77 felépítése, alkalmazási lehetőségek

2.3.1 A rendszer célja, felépítése. A SIS77 egy önálló, pillanatnyilag más adatbázisokhoz nem kapcsolódó általános statisztikai adatfeldolgozó rendszer. Jelenleg elsődleges célja a kórházi morbiditási adatok feldolgozása, de elvileg bármely más statisztikai adatfelvétel feldolgozására is használható. A rendszer több, egymástól többé-kevésbé független programból áll. Ezek a programok különféleképpen kapcsolhatók össze, így különböző feldolgozási folyamatok építhetők fel.

A rendszerműködést egy célnyelv segítségével irányíthatjuk. Ez a célnyelv az operációs rendszer /a job control nyelv/ vezérkártyáiból és az egyes programok paraméterkártyáiból áll, és ilyen módon függ a felhasznált számológéptől illetve az operációs rendszertől. Ebben a vonatkozásban eltér a magasszintű programozási nyelvektől, melyek általában géptől függetlenül definiáltak. Mivel egy speciális célnyelvről van szó, ezért viszonylag sokféle utasítás /paraméterezési forma/ van. A rendszerműködtetés részletes leírására külön tanulmányban kerül sor.

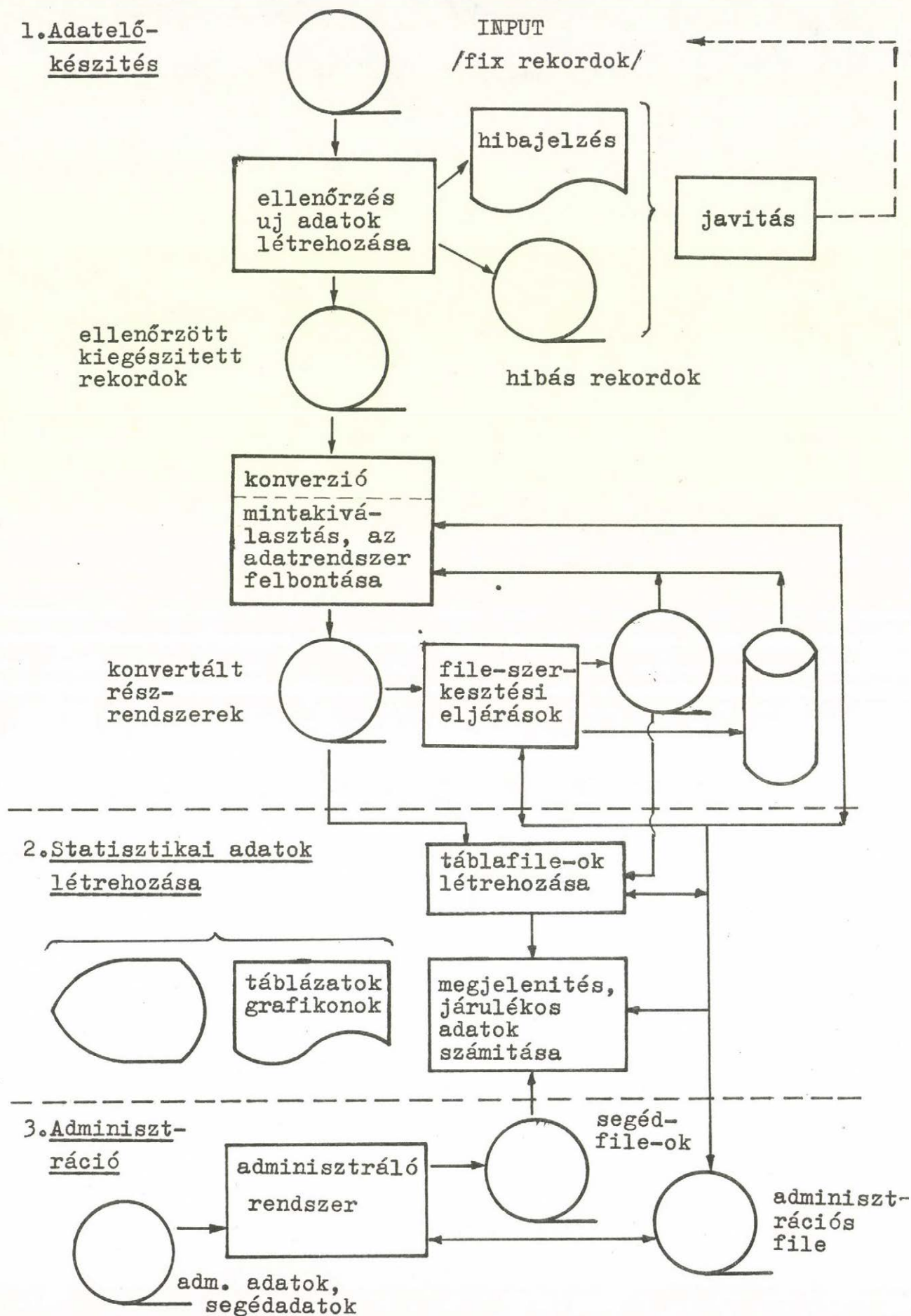
A rendszer működésében három nagy feladatkör választható szét:

- 1.adatelőkészítés /ellenőrzés, kódolás, válogatás, stb/,
- 2.az információszolgáltatás /statisztikai adatok, táblázatok létrehozásának/ folyamata,
- 3.a rendszerműködés adminisztrációja, kiegészítő feladatok.

Ezek a feladatkörök a következő részfeladatokból épülnek fel /ld. pl. [32], [33] /:

- 1.adatellenőrzés, új adatok létrehozása,
- 2.mintakiválasztás, az adatrendszer felbontása /válogatás/,
- 3.konverzió,
- 4.a konvertált fileok átszerkesztése - további adatválogatás, új rekordtipusok kialakítása /pl. kísérőbetegségek, többször ápoltság személyek rekordjai/,
- 5.statisztikai adatok képzése az egyedi adatokból /a táblafájl-ok létrehozása/,
- 6.adatmegjelenítés /táblázatok/, járulékos értékek, statisztikai jellemzők számítása,
- 7.a rendszerműködés adminisztrációja,
- 8.a táblázatok kiírásánál felhasznált szövegek, statisztikai alapadatok bevitele, tárolása.

Ezt a rendszerszerkezetet mutatja be a 3. ábra /ez az ábra tulajdonképpen a 2. ábra egy módosított változata/.



3. ábra

Az ábra részletes elemzésére a következő /2.3.2 - 2.3.9/ pontokban kerül sor.

Az itt felvázolt rendszer a magyarországi kórházi morbiditási vizsgálatokban való sikeres alkalmazásának tapasztalatai alapján egy általános feltételek mellett is optimálisan működő eszköznek bizonyult. Mi biztosítja a hatékony működést? Hogyan valósítja meg rendszerünk a 2.1 és 2.2 pontban felvázolt elveket? Hogyan elégíti ki a rendszer az 1. fejezetben felsorolt igényeket? Ezekkel a kérdésekkel részben egy külön tanulmányban foglalkozunk, részben a következő /2.3.2 - 2.3.9/ pontokban térünk ki rájuk. Előbb azonban röviden összefoglaljuk a megvalósított elvek közül a legfontosabbakat /ld. [33]/:

1. a rendszer teljessége /ld. az előbb felsorolt 1-8. feladatköröket, illetve a 3. ábrát/,
2. az egyes részrendszerek általános formában, paraméterekkel vezérelve működnek,
3. a részrendszerek önállóan is használhatók /modularitás/,
4. az egyes részrendszerek változatos kapcsolódási lehetősége /variálhatóság/,
5. új részrendszerek is beépíthetők /nyíltság/,
6. a felhasznált file-típusok egységesítése /ez a rendszermodulok illeszthetőségéhez kell/,

7. a rendszerműködtetés részletes ellenőrzése, hibajelzések /üzembiztonság/,
8. kényelmes adatbeviteli formák biztosítása,
9. a rendszerben legnagyobb súllyal szereplő feladatok optimalizálása gépidőben, tárolókapacitásban, felhasználói és programozási munkában.

2.3.2 Ellenőrzés, új adatok létrehozása. Az adatellenőrzésben és az új adatok létrehozásában /az átkódolásban/ a SIS77 lényegében háromféle lehetőséget biztosít:

1. megvizsgálható, hogy a kijelölt adatok nem lépik-e túl a megengedett értékhatárokat /korszerű adatrögzítési technikáknál ez az ellenőrzés már az adatrögzítés fázisában is megtörténhet/,

2. a felhasználó által leírt logikai kapcsolatok /függvények/ alapján vizsgálható egyes adatok, adatsorok helyessége, és ugyanígy egy vagy többváltozós függvények leírásával új adatok /kódok/ is definiálhatók,

3. és lehetőség van arra is, hogy a felhasználó tetszőleges - FORTRAN nyelvű - programrészleteket illesszen a rendszerbe /ellenőrzés vagy átkódolás céljából/.

A következőkben a 2. és 3. lehetőséggel foglalkozunk.

Milyen feladatokat kell megoldani összetett ellenőrzési és kódolási feltételek megfogalmazásakor, és számítástechnikailag használható formában történő leírásakor? A SIS77-ben egy olyan módszert alkalmaztunk, amely ezeket a feladatokat egységesen kezeli, egy gyors és biztonságos feldolgozásmódot tesz lehetővé, miközben a felhasználó kényelmét is szem előtt tartottuk.

Az ellenőrzési és kódolási feladatok egy vagy többváltozós függvények kiszámítását jelentik. Ellenőrzés esetén a függvényértékek a "hibás" és "elfogadható" "értékek" lehetnek, új adatok létrehozásakor pedig az új kódértékek. Többváltozós függvények kényelmes és biztonságos leírását, és viszonylag gyors kiszámítását kell tehát megoldani. Ha a szóbanforgó függvény zárt alakban könnyen leírható, akkor az előbb vázolt 3. megoldásmódot célszerű használni /pl. ha a születés dátumából kívánjuk egy adott időpontban az illető ember életkorát meghatározni, vagy pl. ha egy adatot aszerint tartunk elfogadhatónak, hogy egy adott relációban van egy másik kijelölt adattal vagy sem/. Ha azonban a függvényt csak értéktáblázat segítségével tudjuk definiálni, akkor az egyetlen lehetséges megoldás a megfelelő értéktáblázat leírása. Meg kell jegyezni, hogy adatfeldolgozási feladatoknál általában korlátos egész értékeket és változókat tartalmazó

függvények szerepelnek, így az értéktáblázatok megadása elvileg nem okoz problémát. Az értéktáblázatok alkalmazása egyben egy gyors számítási folyamatot is eredményez. Többváltozós függvények esetén azonban komoly problémák léphetnek fel, hiszen az egyes változók viszonylag kis terjedelmű értékkészlete esetén is gyakorlatilag kezelhetetlenül nagy táblázatokra van szükség. Jó példa erre a kórházi morbiditási vizsgálatban központi szerepet játszó diagnózis, nem és életkor összeférhetetlenségi vizsgálat. A felsorolt három változó rendre 10000-féle /diagnózis/, 2-féle /nem/ és 5-féle /életkorcsoport/ értéket vehet fel. Az orvosszakértők által megadott bonyolult adatkapcsolat csak értéktáblázat segítségével írható le. A szükséges értéktáblázat mérete $10000 \times 2 \times 5 = 100$ ezer. Ekkora táblázatot sem a gépben tárolni, sem leírni nem lehet. Egy hierarchikus gráfstruktúra alkalmazásával /ld. [36]/, és egy kényelmes és biztonságos adatbeviteli technika kialakításával /ld. [32]/ tettük lehetővé ennek a problémának a megoldását. Módszerünk tulajdonképpen egy gyors keresési és beillesztési technikát is biztosít. Ezek az utóbb említett feladatok a számítástechnika egy jelentős önálló fejezetét képviselik /ld. pl. [20]/. A módszer részletesebb leírását egy következő tanulmányunkban adjuk, amely elsősorban a SIS77-el kapcsolatos számítástechnikai kérdésekkel foglalkozik.

Meg kell még jegyezni, hogy a SIS77-nek ebben a részrendszerében a fent említett hierarchikus gráf-
strukturáról részletes diagnózis készül, biztonságossá
téve ezzel a felhasználást.

2.3.3 Mintakiválasztás, az adatrendszer felbontása.

A feldolgozandó adatrendszert több okból is részekre
bonthatjuk.

A vizsgált egyedek csoportjainak szétválasztását
egyrészt azért végezzük el, hogy például a további
feldolgozásokban egy adott nagyságu mintával, vagy
csak bizonyos tulajdonságokkal rendelkező egyedekkel
/pl. csak a magyar állampolgárságu kórházi ápoltak-
kal/ foglalkozhassunk. Az adatrendszer ilyen jellegű
felbontása egyrészt a minta reprezentativitásával
kapcsolatos statisztikai kérdéseket vet fel /ld. [23] /,
másrészt bonyolultabb kiválasztási feltételek esetén
matematikai logikai és optimalizálási feladatokat is
meg kell oldani / [29], [30] /. A SIS77 a legbonyolul-
tabb logikai feltételek mellett is egy viszonylag
kényelmesen használható és gyors feldolgozást bizto-
sító válogatási lehetőséget tartalmaz / [29], [32] /.
A kórházi morbiditási vizsgálatban például könnyen
kiválaszthatjuk - mondjuk egy speciális vizsgálat
céljára - azokat a Somogy megyei városokban lakó,

mezőgazdaságban fizikai munkát végző, 40 és 60 év közötti férfiakat, akiket valamilyen szivbetegséggel ápoltak /beleértve a kísérő és következményes betegségeket is/. A SIS77-nek ez a részrendszere megfelelő előkészítés után, direkt elérésű fileok alkalmazásával, egy rendkívül gyors válogatási lehetőséget is biztosít.

Az adatrendszer egy más jellegű szétválasztása abból adódik, hogy bizonyos adatokra egy vizsgálat során egyáltalán nincs szükség. Például a jelenlegi /1972-77. évi/ kórházi morbiditási vizsgálatokban a beteg születésnapjára csak a mintakiválasztásnál van szükség. A rekord egyes adatainak elhagyása nem jelent különösebb számítástechnikai problémát, de az alábbiakban vázolt szempontok érdekes szervezési, optimalizálási kérdéseket vetnek fel. Miről is van szó?

Statisztikai adatfeldolgozásnál a feladatok nagy többsége úgy csoportosítható, hogy egy-egy nagyobb feladatkör csupán néhány adat /sokféle szempontból történő/ vizsgálatát tartalmazza. Így van ez a kórházi morbiditási vizsgálatnál is /ld. [26]/. Ha például egyes vidékek ellátottságát vizsgáljuk, akkor a lakóhely, a kórházkód esetleg egy diagnóziskód mellett más adatokkal nemigen kell törődnünk. Ilyen módon adatrendszerünket több részre bonthatjuk úgy,

hogy egy részrendszerben az egyes rekordokból csak néhány adatot tartunk meg. Ezzel a feldolgozás további részeiben egyrészt sok felesleges adatmozgatást takarítunk meg, másrészt a feldolgozások tárolókapacitás igénye is kisebb lesz.

Az adatrendszer ilyen típusu felbontásánál meg kell találni az optimális középutat. Nem szabad túl sok részrendszert létrehozni, mert a válogatási feladatok több erőforrást lekötnek mint amennyit a kisebb részadatrendszerek alkalmazásával nyerünk, és a sok kisméretű részrendszer végülis együttesen tulságosan nagy tárolóhelyet foglal el. Olyan részrendszereket kell létrehozni amelyek viszonylag kevés adatot tartalmaznak, mégis a lehető legtöbb kérdésre választ lehet belőlük nyerni. Ehhez természetesen előre ismerni kell az igényelt információkat /a várható információigényeket/, és előre meg kell tervezni a teljes adatrendszer részrendszerekre való bontását. Ez a feladatkör érdekes kombinatorikai, /sztochasztikus/ optimalizálási problémákat rejt magában.

2.3.4 Az alapadatforma konverziója. A SIS77 rendszerben különös gondot fordítottunk az optimális megoldásmódok keresésére és alkalmazására. Ezt tartottuk szem előtt akkor is, amikor a karakter formában tá-

rolt alapadatok konverzióját megvalósítottuk. Az adatok tárolásával kapcsolatban a következő probléma merül fel: Karakter formájú adatok viszonylag kis helyet /BCD kódban karakterenként 6 bitet/ foglalnak el, viszont a számítási eljárások lassubbak. Bináris tárolás gyors feldolgozást tesz lehetővé, de nagy a helyigénye /pl. a HWB 66-os gépeken 36 bit/.

Mivel statisztikai adatfeldolgozásnál kódolásra általában természetes számokat használunk, ezért lehetővé vált az, hogy az adatokat tömören, és ugyanakkor a gyorsan feldolgozható bináris formában helyezhessük el. Egy-egy gépi szóban egyszerre több adatot is tárolunk. Egy adat annyi bitet foglal el, ahány az illető adat maximális értékének tárolására elegendő. A bitenkénti finomságban történő tárolást a Honeywell FORTRAN FLD függvénye tette lehetővé. Más rendszerekben, ahol FORTRAN nyelven bitek kezelésére szolgáló eljárás nincs, ott egy néhány soros assembler program beépítésével oldható meg ez a feladat. Előjelproblémák az előbb említett ok /a kódok természetes számok/ miatt nincsenek. Ez a tárolásmód egyébként még a karakteres tárolásnál is tömörebb. Például egy kétjegyű szám, mondjuk az ápoltság születésének hónapja két karakternyi, azaz tizenkét bitnyi helyet foglal el. Binárisan tárolva

ugyanaz az érték négy biten is elhelyezhető.

2.3.5 Speciális file-ok, file szerkesztési eljárások.

Statisztikai adatok létrehozása /ld. táblafile-ok/
és újabb válogatási eljárások előtt sokszor célszerű
a rendelkezésünkre álló válogatott, konvertált file-
okat módosítani /átszerkesztteni/. A SIS77 a követke-
ző lehetőségeket tartalmazza:

2.3.5.1 A feldolgozandó rekordok rendezése lehetővé
teszi nagyméretű táblafile-ok gyors létrehozását,
és a gyors válogatási eljárás is rendezett file-okból
képzett direkt elérésű file-okon működik.

2.3.5.2 Ha egy feldolgozásban a vizsgált rekord né-
hány adatát nem tudjuk megkülönböztetni, - mint pl.
a kórházi morbiditási vizsgálatban a felsorolt kísérő
és következményes betegségek is egyenrangúak -
akkor a rekordokat ezek szerint az adatok szerint meg-
kell többszörözni, hogy olyan rekordokat hozzunk lét-
re, amelyekben minden adat szerepe egyértelmű /pél-
dául ápolási esetenként annyi új rekordot hozunk lét-
re, ahány kísérőbetegség szerepel az eredeti rekord-
ban/.

2.3.5.3 Más esetekben több rekord is tartozhat egyet-
len egyedhez - pl. a többszörösen ápolat személyek

esetében. Ilyenkor különböző rekordok /pl. különböző ápolási esetek/ összevonását kell elvégezni.

A két utóbbi kérdéskör /2.3.5.2 és 2.3.5.3 pont/ az adatbázismodellek problémáihoz - például a relációs adatmodellek témaköréhez /[5]/ - kapcsolódik.

2.3.5.4 Statisztikai vizsgálatokban szükség lehet kijelölt adatpárok előfordulási gyakoriságának vizsgálatára is /például meg akarjuk tudni, hogy mely diagnózisok és milyen gyakorisággal párosulnak a kórházban ápolts személyeknél/. Ilyen célból a SIS77-ben olyan file-t lehet létrehozni, amelyben kijelölt adatok /pl. egyes diagnózisok/ különböző értékpárjaihoz külön rekordok tartoznak.

2.3.6 Táblafile-ok létrehozása. A táblafile-ok - mint az előzőkben /2.1.2.4 pont/ már elmondtuk - tulajdonképpen nagyméretű statisztikai táblázatok. Egy kijelölt kódkombináció különböző értékeinek gyakoriságát /ápolási esetek száma/ és egy másik kód összegértékeit /ápolási idők összege/ tartalmazzák, legfeljebb nyolc szempont szerinti bontásban. A táblafile-ok alkalmazása - ugyanugy mint a SIS77 sok más elemének létrehozása - a feldolgozások optimalizálásának érdekében történt. Táblafile-okat /kezelhető formában/ tulajdonképpen már az 1972-73.

évi adatfelvétel feldolgozásakor is használtunk / [21], [26] /, további kísérleteinkben viszont a táblafile már központi szerepet játszott / [11], [22] /.

Miért előnyös a táblafile-ok alkalmazása? Miért nem közvetlenül az egyedi adatokat /rekordokat/ tartalmazó file-okból képezzük a szükséges statisztikai táblázatokat? Válaszként a következő szempontokat sorolhatjuk fel:

2.3.6.1 Az egyedi adatok halmazából viszonylag nagy gépidőfelhasználással hozhatjuk létre a statisztikai adatokat, még akkor is, ha az előzőekben tárgyalt /ld. 2.2 pont/ optimalizálási módszerek alkalmazásával többszörösére növeljük a számítási sebességet /ld. [34], [35] /. Például a jelenlegi /1974-77.évi/ 150 ezres mintanagyságu kórházi morbiditási vizsgálatnál ezek a számítások néhány percnyi /kb. 1-2 perc/ időt vesznek igénybe, míg a táblafile-okból készített statisztikák létrehozásához a mintanagyságtól függetlenül, néhány másodperc elegendő. Ezek az időadatok az ÁSZSZ HWB 66/60-as gépére vonatkoznak. A futásidő különbség oka a következő: az egyedi adatok file-ja általában nagyobb mint a belőle létrehozott táblafile-ok - esetünkben a mintegy 150 ezer kórházi ápoltságait tartalmazó válogatott és konvertált részrendszerek néhány százezer szónyi tárolóterületet foglal-

nak el, a táblafíle-ok viszont csak néhányszor tízezer szót. A méretkülönbségek mellett az időkülönbség másik oka az, hogy az egyedi adatokból hosszadalmasabb számításokkal nyerhetők a kívánt statisztikai adatok, mint a már eleve statisztikai adatokat tartalmazó táblafíle-ból. Ez még akkor is lényeges sebességkülönbséget biztosít, ha a táblafíle mérete megközelíti, vagy akár kissé túl is haladja az eredeti file méretét.

2.3.6.2 Figyelembevée a futásidőkre tett előző megállapításainkat, a táblafíle-ok alkalmazása minden további meggondolás nélkül is célszerűnek mondható, ha csak azt vesszük figyelembe, hogy egyes statisztikai táblázatok előállítására ismételten sor kerülhet /pl. a megrendelők újabb példányokat, vagy valami apró formai módosítást igényelnek/.

2.3.6.3 A táblafíle-ok alkalmazásának fő előnye azonban a variálási lehetőségekben rejlik. A SIS77-ben jelenleg legfeljebb nyolc szempont szerinti bontásban lehet táblafíle-t készíteni, és ebből legfeljebb négydimenziós statisztikai táblázatok állíthatunk elő. Még ha ezt a megszorítást figyelembe is vesszük, akkor is lehetséges $\binom{8}{4} + \binom{8}{3} + \binom{8}{2} + \binom{8}{1} = 162$ különböző táblázat létrehozása.

2.3.6.4 Tovább növekszik a variációs lehetőségek száma, hogyha figyelembe vesszük a tábladimenziók /bontási szempontok/ sorrendváltozását is. Ekkor a létrehozható táblaváltozatok maximális száma 2080. Természetesen ez csak egy lehetőség, amelyen belül a felhasználó csak néhány táblatípust igényel.

Mindezek mellett azonban még olyan variációs lehetőségek is vannak, amelyek a különböző táblaváltozatok megszámlálhatatlan sokaságát adhatják. A táblafile-nak a megjelenítésre kiválasztott dimenzióin belül tetszőleges összevonásokat /pl. életkorcsoportok/, kihagyásokat, részösszegképzéseket /pl. az összes vidéki ápoló összevonása/ alkalmazhatunk. Készíthetünk táblázatokat kumulatív vagy nem kumulatív formában, grafikonnal vagy anélkül. Az egyéb formai változtatások mellett a táblázatban megjelenő viszonyszámokat is módosíthatjuk, cserélhetjük /pl. 1000 lakosra jutó ápolási eset, 10 ezer lakosra jutó ápolási eset, egy kórházi ágyra jutó ápolási eset, stb/.

2.3.6.5 A kritikus olvasóban felmerülhet azonban az a kérdés, hogy vajon a táblafile-ok dimenziószámának növekedésével nem növekszik-e méretük olyan rohamosan, hogy az eddig felsorolt előnyök valójában nem használhatók ki. Például egy nyolcdimenziós táblafile-nál,

még ha az egyes dimenziók csak tízféle értéket vehetnek fel, akkor is egy százmillió elemű táblafíle-t kapunk, amelynek a létrehozása, a tárolása és a felhasználása egyaránt lehetetlen. Erre a megjegyzésre a következő ellenérvek hozhatók fel /példaként most is a kórházi morbiditási vizsgálatra hivatkozunk/:

A statisztikai táblázatok egy jelentős részében olyan adatok szerepelnek, amelyek két-háromféle értéket vehetnek csak fel. Ilyenek például a beteg neme /férfi, nő/, a lakóhely típusa /főváros, megyei város, város, község/, a beutalás módja /sürgős, nem sürgős/, a halálozás jelzése /a beteg meghalt, nem halt meg/, a kórház típusa /oktatási intézmény vagy sem/, stb.

Egy másik ok amely a dimenzióméretetek csökkenéséhez vezet az, hogy egyes adatoknak csak valamilyen összevont kódjára van szükség /pl. életkorcsoportok, ápolási időcsoportok, stb/. Ha az egyik dimenzió igen részletes bontásban szerepel /pl. a három vagy négyjegyű betegségkód/, akkor más dimenziók eleve nem kerülhetnek részletes bontásban megjelenítésre.

Bizonyos témakörök, a táblafíle-ban együttesen szereplő adatok sokszor automatikusan kizárják az egyes bontási szempontok finomításának szükségességét. Például a beteg foglalkozásával kapcsolatos

kérdéseknél az életkor gyermekeknél történő részletes bontására nincs szükség, ugyanakkor mondjuk egyes diagnóziscsoportok vizsgálatakor csak a csecsemőkoru ápoltak adatai az érdekesek, tehát különböző felnőtt korcsoportokat fölösleges létrehozni.

2.3.7 Az adatok /eredmények/ megjelenítése. A SIS77 működésének elsődleges célja statisztikai táblázatok készítése. Ezeket a táblázatok a rendszer mindig valamilyen táblafile-ból hozza létre.

A SIS77 táblázatkészítő részrendszere a következő lehetőségeket biztosítja a felhasználó számára.

A táblázatok gyakoriságértékeket és egy kijelölt kód összegértékeit tartalmazzák /ápolási eset-szám és ápolási napok összege/. Ezek az értékek közvetlenül a táblafile-ból nyerhetők. A táblázat egyik bontási szempontján belül százalékos megoszlás is kiírásra kerül, mind a gyakoriságra, mind a kódösszegre, és a két érték hányadosa /átlagos ápolási nap/ is megjelenik. Ezek az adatok minden táblázatban megtalálhatók. A felhasználó kívánságára statisztikai viszonyyszámokat /10 ezer lakosra jutó ápolási eset, stb/ is tartalmazhat a táblázat.

A táblafile-okból tetszőleges bontási szempontokat kiválasztva készíthetünk legfeljebb négydimenziós táblázatokat. Az egyes dimenziókon belül tetsző-

leges összevonásokat képezhetünk. Ha például a tábláfile-ban a betegségkód az "A" jegyzék szerinti bontásban szerepel /ez egy 150 elemű kódlista/, akkor a kinyomtatott táblázatban az "A" jegyzéken kívül lehet a 17 tagú betegségfőcsoport beosztást is használni, stb. Egy kijelölt bontási szemponton belül gyakorlatilag tetszőleges módon képezhetünk részösszegeket. Erre sok esetben szükség van - a különbségek vagy hasonlóságok kiemelése érdekében célszerű egymás mellett szerepeltetni a részletesen bontott és az összevont értékeket /pl. a mezőgazdaságban dolgozó szellemi foglalkozású és az összes szellemi foglalkozású kórházi ápolási adatait/.

A felhasználó a gyakoriságokról és a kódösszegekről grafikont is készíthet. Mivel a grafikus megjelenítés célja a különböző értékek gyors összehasonlíthatóságának biztosítása, ezért összehasonlítható értékeket kell ábrázolni. Erre a célra a statisztikai alapadatokkal /pl. népességstatisztikai adatokkal/ osztott értékek alkalmasak /pl. 10 ezer lakosra jutó ápolási eset és nap/.

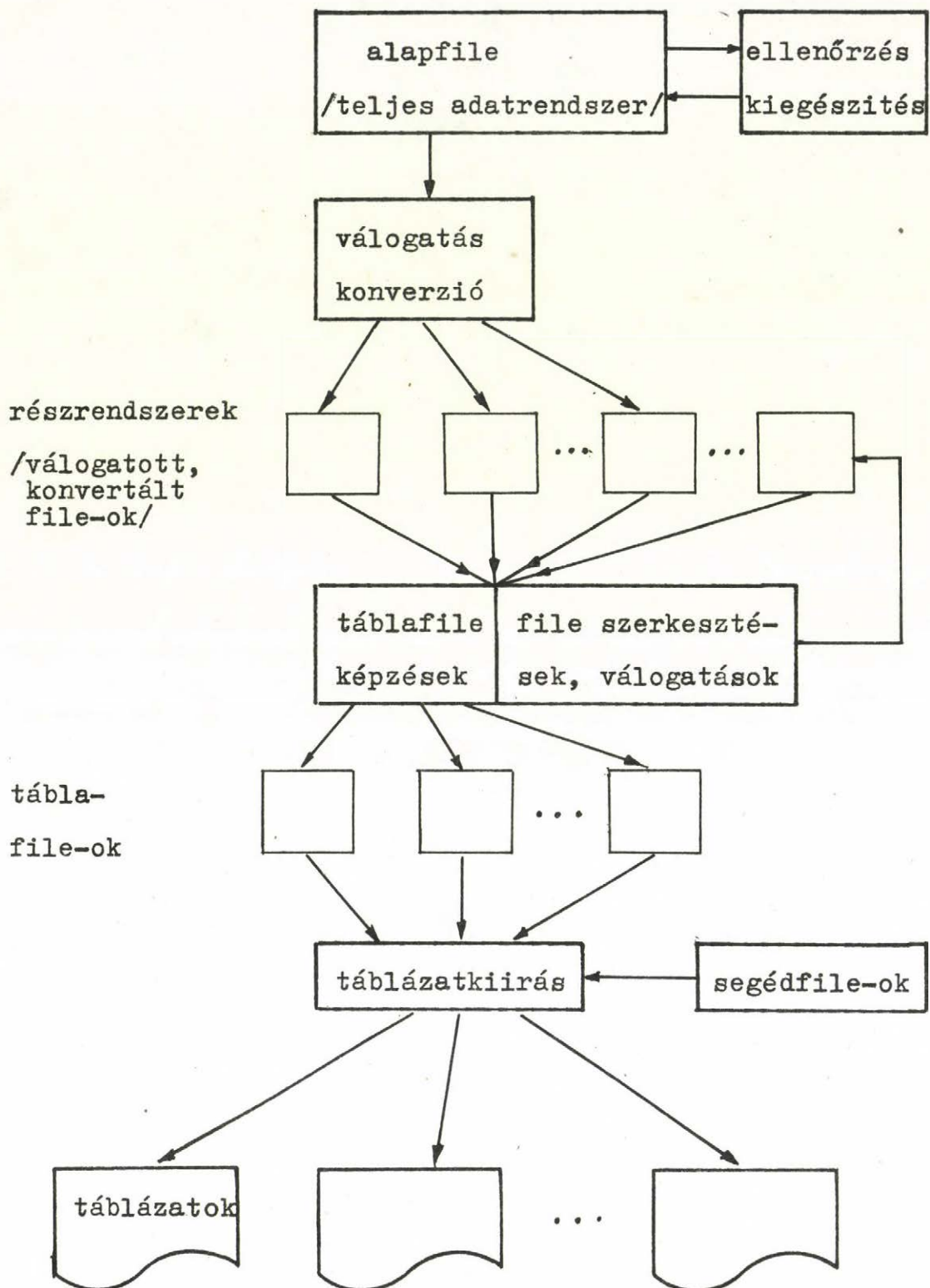
A részösszegek /ugyanúgy mint a teljes összegek/, a kumulatív értékek megadása, grafikus megjelenítés, százalékszámítás mindig csak egy kijelölt bontási szempontra igényelhető. Ez a bontási szempont a kiírt táblázat legbelső dimenziója /a legalacsonyabbrendű

bontási szempontja/ lehet csak.

Ki kell emelni, hogy az előzőkben felsorolt lehetőségek általános statisztikai adatfeldolgozási igényeket elégítenek ki. A kódösszegekkel kapcsolatban még annyit megemlíthetünk, hogy azok nemcsak átlagszámításra /a kódösszeg és a gyakoriság hányadosa/, hanem más additív statisztika számítására is lehetőséget adnak. Ha ugyanis a feldolgozás első lépésében /ld. 2.3.2 pont/ kódérték szorzatokat és négyzetértékeket képezünk, akkor táblázáskor szorzat és négyzetösszegeket kaphatunk. Ezeknek az értékeknek a felhasználásával statisztikai analízisek egy széles köre - a szokásos többváltozós analízisek, mint pl. szórásanalízis, stb - elvégezhető.

A táblázatok megjelenítésére két részrendszer szolgál. Az egyik a kötegetelt /batch/ feldolgozások céljára, a másik - bizonyos korlátozásokkal - TSS-ben való futtatásra, és képernyőn való megjelenítésre szolgál.

Mindkét részrendszerben a táblázatok fejlécei, az egyes adatok, adatértékek megnevezései a felhasználó által tetszőleges módon változtathatók. Batch üzemmódban nagytömegű adatmegnevezés esetén /pl. a betegségmegnevezéseknél/ illetve gyakran használt adatok

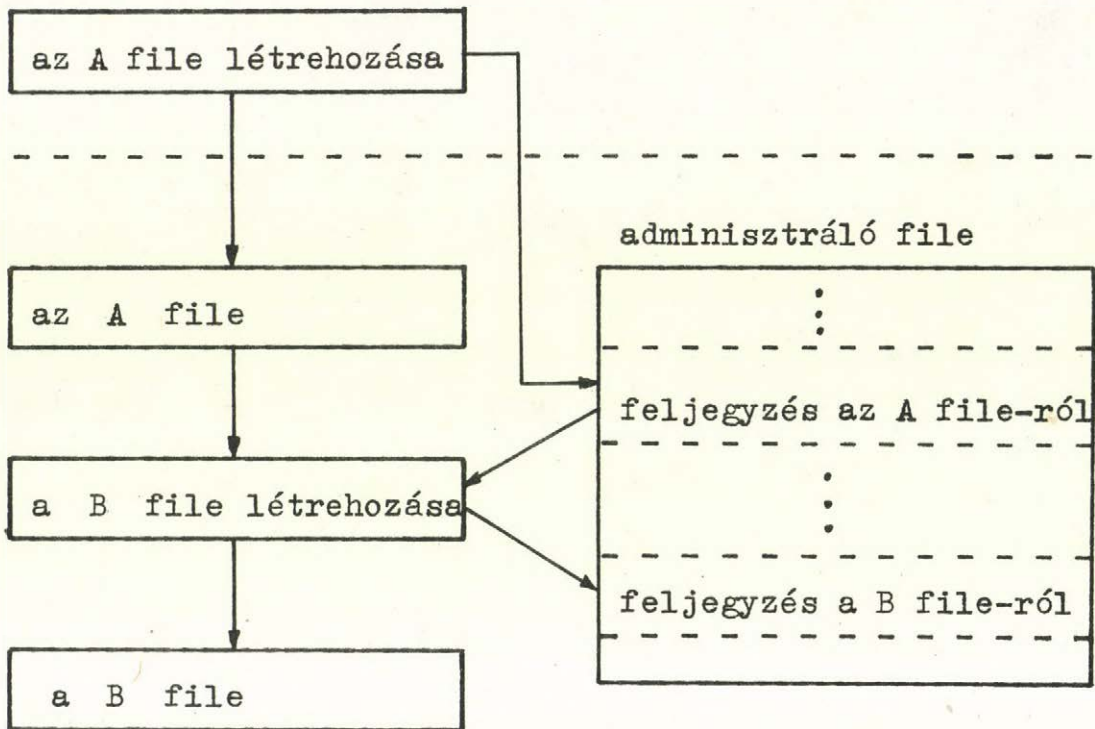


4. ábra

megjelenítések /pl. megyék, kórházi szakmák/ a rendszerben tárolhatók, és a kívánt esetben kiírhatók az aktuális szövegek /a felhasználónak nem kell ezeket újra és újra leírnia vagy lyukkártyán tárolnia/.

2.3.8 A rendszeradminisztráció. A rendszeradminisztráció szükségességét, és egyben a SIS77 alkalmazásának előnyeit a 4. ábrán bemutatott file-kapcsolat vázlattal érzékeltethetjük.

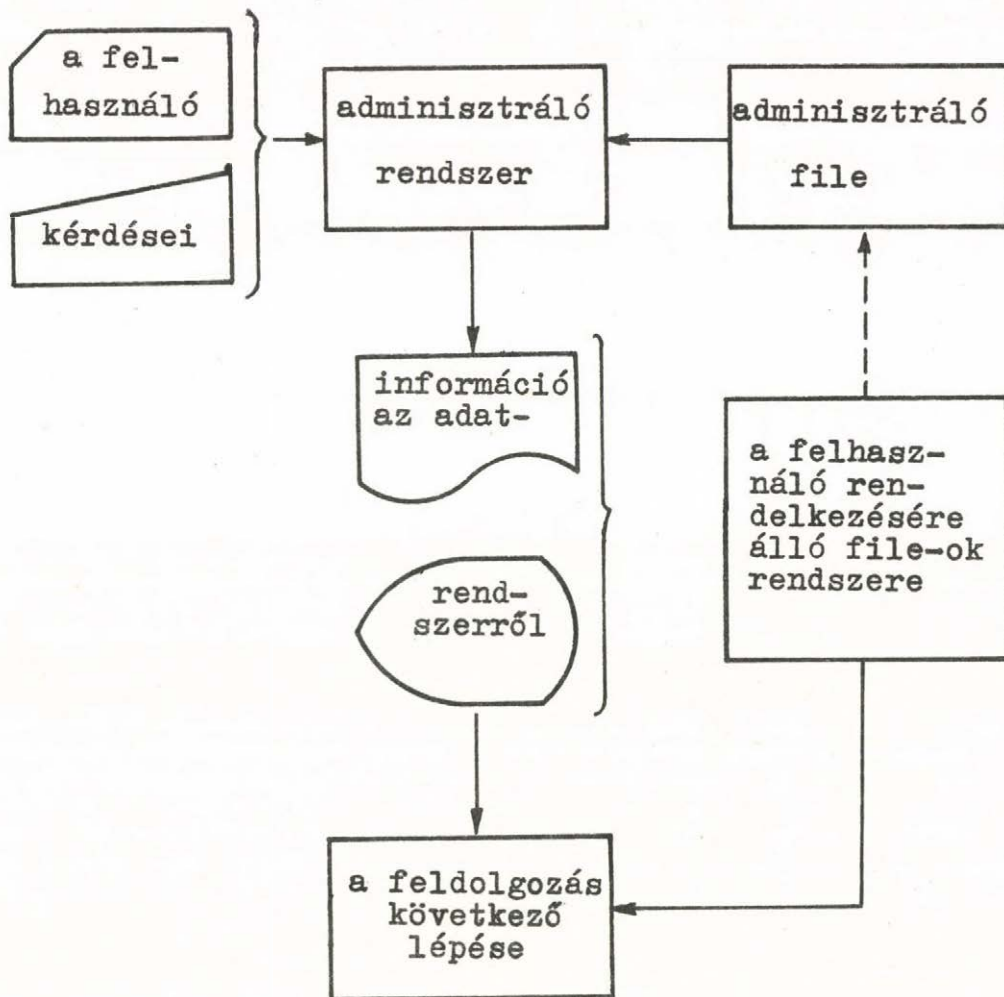
A feldolgozási folyamat során a válogatások, file-szerkesztések, táblafile-képzések során file-ok egy szerteágazó sokaságához jutunk /ld. 4. ábra/. Egy táblázatot tehát file-átalakítások sorozatának végeredményeként kapunk /éppen ebben rejlik a rendszer egyik nagy előnye, a variálhatóság/. Meg kell azonban jegyezni, hogy az így kapott file-sorozatok kapcsolatának ábrázolásával általában nem egy fa-szerkezetű gráfhoz jutunk, mert az egyes átalakítás-sorozatok találkozhatnak is /esetenként egy-egy táblázatot többféle táblafile-ból is előállíthatunk, stb./. A feldolgozási folyamatok - még ha például optimális megoldásokra törekszünk - nem feltétlenül egyértelműek. Mindenesetre egy nagyobb feldolgozási rendszer esetén file-ok nehezen áttekinthető sokasága jön létre. Ezért szükséges minden szinten adminisztrálni a létrehozott file-okat /ld. 3. és 5. ábra/. Ezt a felada-



5. ábra

tot /néhány speciális kivételtől eltekintve/ a SIS77 automatikusan elvégzi.

Az adminisztráció másik előnye a nagyfoku biztonság. A rendszer által rögzített jellemzők a feldolgozás következő lépéseiben hozzáférhetők, így a tévedés kizárt. Ezt a kapcsolatrendszert szemlélteti az 5. ábra. Az ábrán bemutatott feldolgozási lépésben egy A file-ből hozunk létre egy B file-t. Az A file jellemzőit a megfelelő rendszerprogram egy előző lépésben már rögzítette az adminisztráló file-on.



6. ábra

Az adminisztráció lehetővé teszi azt, hogy a feldolgozás közbeeső lépéseiben a felhasználó tájékoztatást kapjon a rendszer pillanatnyi állapotáról és a továbblépési lehetőségekről. Ezt a feldolgozási lépést mutatjuk be a 6. ábrán. A felhasználó információkat kér a rendszer pillanatnyi állapotáról, és ennek megfelelően teszi meg a feldolgozás következő lépését.

2.3.9 Segédfile-ok. Statisztikai táblázatok készítésekor az alapadatokon kívül számtalan kiegészítő adatra van szükség.

A táblázatok szöveges részében felhasznált adatmegnevezések /pl. betegségnevek/ néha olyan tömegben szerepelnek, hogy célszerű számukra a gépen tárolóhelyet fenntartani. Ugyanigy a statisztikai viszonyszámokat /pl. a népességstatisztikai adatokat, a kórházi ágyak számát, stb/ is rögzíthetjük valamilyen háttértárolón. Az adminisztráló részrendszer ezeket a file-okat ugyanigy számontartja, mint a feldolgozandó adatrendszer file-jait, és feltöltésükről is az adminisztráló rendszer gondoskodik.

A segédfile-ok, az adminisztráló rendszer és az adminisztráló file - ugyanigy mint a file-szerkesztési feladatok /ld. 2. ábra/ - beilleszthetők egy adatbázisba is, amelyhez részként csatlakozik a statisztikai rendszer. Ennek a tanulmánynak azonban nem célja, hogy komplex adatbázisok felépítésével foglalkozzék, így ezeknek a kérdéseknek a részletezésére most nem térünk ki.

2.3.10 Kapcsolatok más rendszerekkel. Befejezésként néhány szót szólunk a SIS77 más rendszerekhez való kapcsolódásának lehetőségeiről, a SIS77 programjainak önálló felhasználásáról.

Az ellenőrző, kiegészítő részrendszer és a mintakiválasztás, mintanagyság ellenőrzés feladatai gyakorlatilag tetszőleges adatfeldolgozó rendszerenél alkalmazhatók. Ugyanígy bármely más adatfeldolgozó rendszerben előkészített /ellenőrzött/ fix rekordu szekvenciális file-ok statisztikai vizsgálata elvégezhető a válogatás, a táblafile-készítés és táblázatkiírás részrendszereivel.

A válogató rendszer - bár outputja speciális, tömörített bináris formájú - egy konvertáló program közbeiktatásával, nem csak statisztikai vizsgálatoknál, hanem bármely más rendszerben képes a bonyolult feltételekkel leírt adatrendszerfelbontási feladatokat elvégezni.

A táblázatkiíró rendszer többdimenziós adatmátrixok változatos igények szerinti megjelenítésére alkalmas - ld. 2.3.7 pont.

További alkalmazási lehetőségek a rendszer különböző irányokban történő kiterjesztésével érhetők el. Erről a kérdéskörrel szól a következő 3. fejezet.

3. A SIS77 továbbfejlesztési lehetőségei

A rendszer alkalmazási hátterét bemutató 1. fejezetben már szóltunk a jelenlegi kórházi morbiditási adatfeldolgozásban folyó vizsgálatok kiterjesztési lehetőségeiről. Most a rendszer továbbfejlesztésének számítástechnikai kérdéseivel foglalkozunk, természetesen állandóan szem előtt tartva az alkalmazási igényeket.

A továbbfejlesztési lehetőségeket két csoportban tárgyaljuk. Először a felhasználót közvetlenül érintő kiegészítési javaslatainkat soroljuk fel. Ezután térünk rá a rendszer belső felépítését érintő kérdések tárgyalására, amelyek egyrészt szervezői problémákkal, másrészt a rendszer működésének hatékonyságával vannak kapcsolatban.

A továbbfejlesztések alapját mindig az az általános elv képezi, hogy a rendszer egyes részeihez nem speciális, részfeladatokat megoldó programokat /programrészeket/ illesztünk, hanem - az előző fejezetekben megfogalmazott igényekhez és elvekhez igazodva - az egyes részrendszereket úgy alakítjuk át és olyan új részrendszereket hozunk létre, hogy a teljes rendszer minél szélesebb feladatkör minél hatékonyabb megoldására legyen képes.

3.1 Az alkalmazás kiterjesztése

3.1.1 Adatelőkészítés, az adattartalom bővítése, áttérés más számítógépekre. Ebben a pontban a rendszer általános használhatóságával kapcsolatos lehetőségekkel foglalkozunk.

3.1.1.1 A kórházi morbiditási rendszerben - de más adatfeldolgozó rendszerek esetén is - az adatfelvétellel sokszor egy rendkívül munkaigényes kódolási eljárást /pl. diagnózis, foglalkozás, lakóhely kódolása/ is szükségessé tesz. Ez a feladat általában magas képzettségű munkaerőt - orvosok, szervezők - igényel. Ezért célszerű a kódolást automatizálni valamilyen szövegkezelő, szövegfelismerő részrendszer beépítésével.

Ebben a kérdéskörben számos irodalmi forrás áll rendelkezésünkre. Egy általános, áttekintő tanulmányként említhetjük a G. Gerbner szerkesztésében megjelent *The analysis of communication content* című cikkgyűjteményt /John Wiley, New York, 1969./, vagy T.J. Stone és társai munkáját: *The general inquirer, A computer approach to content analysis* /MIT, Cambridge, Massachusetts, 1966./. Az utóbbi az általános módszerek és elvek mellett alkalmazási példákat is bemutat. Egy a hatvanas évek eredményeit összefoglaló bibliográfia:

Manchin Gy., Tartalomelemzés a szociológiában, Tömegkommunikációs Kutató Központ, 1969. Ujabb elképzeléseket és eredményeket foglal össze: A. Deichsel, Elektronische Inhalts Analyse /Verlag Spiess, Berlin, 1975./ és T.J. Stone, Standards for computer aided content analysis: the Pisa conventions and recommendations /Social Science Informations, 1975/1. /.

Ismételten felhívjuk a figyelmet arra, hogy a szöveg alapján történő automatikus kódolás problémáinak megoldásával egy nem csak a hospitalizált morbiditás vizsgálatában, hanem más adatrendszerek feldolgozásánál is használható eszközhöz juthatunk.

3.1.1.2 Olyan, rendszeresen ismétlődő adatfelvételek vizsgálatánál, mint amilyen a kórházi morbiditási rendszer is, célszerű lehetővé tenni egy alkalmanként változó kérdéscsoport beiktatását.

Egyik évben például a szivbetegség adataival, másik évben mondjuk a mezőgazdasági dolgozók, vagy éppen a nagyvárosi lakosság speciális problémáival lehetne részletesebben foglalkozni. Ugyanigy a teljes populációra vonatkozó kiegészítő vizsgálatokat is el lehet végezni évenként más és más témára, például az életmód /lakáskörülmények, család, munkahelyi közérzet, dohányzási szokások, stb/ és a megbetegedések - és más hasonló adatcsoportok - kapcsolatának tanulmányozásával.

Ezek a célok úgy valósíthatók meg, hogy a minden évben felvett adatok mellett fenntartunk egy korlátozott nagyságu szabad adatterületet.

A SIS77 már jelenlegi formájában is lehetővé tesz ilyen változó tartalmu feldolgozásokat, csak a felhasznált adatlapon kell a szükséges kérdések és kódpozíciók helyét biztosítani.

3.1.1.3 A SIS77 jelenleg egy HWB 66/60-as gépen GCOS operációs rendszerben működik. Azonban már az egészségügyi alkalmazások terén is - nem beszélve egyéb felhasználási lehetőségekről - előtérbe kerül más gépek /elsősorban ESZR rendszerek/ felhasználásának igénye. Mint ahogy az előzőkben már leszögeztük /ld. pl. 2.2.2.1 pont/, a SIS77 átvihető más számítógéprendszerekre is. Ilyen módon lehetővé válik a rendelkezésre álló ESZR /vagy IBM/ gépkapacitások kihasználása, és ennek folytán hazai és nemzetközi kapcsolatok kiépítése. Az utóbbi vonatkozásban /elsősorban a KGST-n belül/ az egészségügyi alkalmazások területén sok lehetőség nyílik - ld. Mányi Géza, Az Egészségügyi Minisztérium Központi Számítástechnikai Intézetének szerepe az irányítási információrendszer részeként működő egységes, országos betegnyilvántartási rendszerek kidolgozásában /Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológi-

ában, 8. Kollokvium, Kerekasztal-megbeszélés, Szeged, 1977./.

3.1.2 Az adatbázis továbbfejlesztése, interface-ek.

Egy adatfeldolgozó rendszer használhatóságát nagymértékben az határozza meg, hogy mennyire bonyolult adatbázisokkal képes dolgozni, és más rendszerek irányában milyen kapcsolódási lehetőségeket biztosít. Ezen a területen a SIS77 következő továbbfejlesztéseit javasoljuk:

3.1.2.1 A hospitalizált morbiditás témakörén belül elsősorban a többször ápoltak vizsgálatánál /több évi beteganyag együttes vizsgálata, egyes ápolts személyek követése/ van szükség adatbázisfejlesztési munkákra. Ez a feladat orvosi, szervezői és számítástechnikai /matematikai/ problémák egyidejű megoldását, és a célok egzakt megfogalmazását igényli.

3.1.2.2 A kórházi morbiditási rendszer az egységes egészségügyi információs rendszer része. Ilyen módon szükséges az általa használt adatbázisnak az egységes rendszerbe /adatbázisba/ történő integrálása. Ez fejlettségében egy ténylegesen integrált adatbázis kiépítését feltételezi /ld. pl. [33] illetve az 1. ábrát/, de alacsonyabb szinten is megvalósítható az egyes részrendszerek közti interface-ek kiépítésével.

Ugyanitt emlithetők a folyamatos, vagy több lépésben végzett adatfelvétel, ellenőrzés, javítás műveleteivel kapcsolatos problémák is.

3.1.2.3 Statisztikai adatfeldolgozásban fontos szerepet játszanak a különböző matematikai statisztikai vizsgálatok. Bár a SIS77 közvetett módon jelenlegi formájában is lehetővé teszi a szokásos statisztikai analízisek egy széles körének alkalmazását /ld. 2.3.7 pont/, mégis célszerűnek látszik a programkönyvtárakban hozzáférhető szubrutinok, programok /pl. az ÁSZSZ Honeywell gépén a BMD programgyűjtemény/ közvetlen alkalmazhatóságát biztosítani. Ez a cél a táblafájl-ok /ld. 2.3.6 pont/ továbbfejlesztésével érhető el.

3.1.2.4 A SIS77 egyes fázisaiban /táblafájl-ok, válogatott, konvertált fájl-ok, speciális fájl-ok/ sajátos tárolásmódot alkalmaztunk. Ha tehát valaki ezekben a pontokban akar hozzáférni a rendszerben tárolt információkhoz, - például egy speciális részpopulációt akar a SIS77-től függetlenül további vizsgálatok céljára kiválasztani - akkor ezt konverziós programok segítségével teheti meg. Célszerű ilyen konverziós programokat előre elkészíteni, és a rendszerbe beilleszteni.

3.1.3 Az input és output formátumok fejlesztése. A felhasználót egyik legjobban érdeklő kérdés az, hogy az eredményeket milyen formában kapja kézhez, és mennyi munkaráfordítással.

3.1.3.1 A SIS77 a felhasznált számológép központi tárcapacitásának függvényében, csak korlátozott nagyságú táblázatok előállítására képes. Nagy táblázatoknál nagy memóriaterületet használ fel, és így ezeknek a táblázatoknak az előállítására nagyobb költségeket is kell fordítani. Ilyen kényszerítő okok miatt a nagyméretű táblázatok /pl. 40-50 ezer sor esetén/ részletekben célszerű előállítani. Ez a megoldásmód azonban a felhasználó számára kényelmetlen. Ezért javasolható a táblázatkiró részrendszer /ld. 2.3.7 pont/ olyan módon történő kiegészítése, hogy nagyméretű táblázatok létrehozását egy lépésben, kis memóriafelhasználással is lehetővé tegye.

3.1.3.2 A táblázatkiró részrendszer jelenleg csak egy táblatípus megjelenítését teszi lehetővé, bár a változatos paraméterezési lehetőségek ezen az egy táblatípuson belül is sokféle felhasználási módot biztosítanak /ld. 2.3.7 pont/. A kórházi morbiditási vizsgálaton belül - és esetleg más alkalmazások kapcsán is - várhatóan felmerülő új formátumigények /pl. a

hagyományos "keresztáblák" kinyomtatásának igénye/ kielégítésének érdekében célszerű a várható új formátumigényeket összegyűjteni, az igények megfogalmazását egy lehetőleg minél egységesebb formába önteni, és megvalósításuk lehetőségét a SIS77 rendszerébe beépíteni. Mivel a felhasználók számára hozzáférhető táblázatkészítési módok /ld. pl. a 2. fejezet bevezetőjében említett könyvtári programokat/ csak igen szegényes választékot kínálnak, ezért az igények teljes körű feltárása - még a hospitalizált morbiditás témakörén belül is - orvosi, szervezői és számítástechnikai szakemberek együttműködését követeli.

3.1.3.3 Bár statisztikai feldolgozásoknál nem elsődleges cél az eredmények interaktív módon, terminálról való elérése, mégis célszerű a modern rendszerek lehetőségeit kihasználva egy gyorsabb, időosztásos /time-sharing/ üzemmódban történő futtatás lehetőségének biztosítása is.

Itt elsősorban a táblázó program és az adminisztráló program /ld. 2.3.8 pont/ interaktív üzemeltetésének továbbfejlesztése kívánatos. Az ehhez szükséges alapok: a nagy sebesség, viszonylag kis memóriaigény, a SIS77-ben adottak.

3.1.3.4 A felhasználó kényelme és a biztonságos üzemeltetés érdekében a jelenlegi számkódolásu vezérlés

helyett érdemes áttérni - a magasszintű nyelvekhez hasonlóan - egy szöveges utasításkészlet alkalmazására. Ez az igény a SIS77 adminisztráló részrendszerében már megvalósult, és az interaktív üzemmódra való áttéréskor /ld. az előző 3.1.3.3 pontot/ szükségessé is válik.

3.2. A szervezés és a belső felépítés kérdései

3.2.1 A szervezést segítő kiegészítések. Az eredményeket felhasználó szakemberek szempontjainak figyelembevétele mellett a rendszert üzemeltető munkaerő jobb kihasználása érdekében célszerű a feldolgozás minden egyes lépésében a rendszerkezelő munkáját is hatékonyabbá tenni /ld. 2.2 pont/.

3.2.1.1 A SIS77 első részrendszerében /ellenőrzés, új adatok létrehozása/ egy nagyon általános lehetőség van az eredeti adatok átalakítására, új adatok létrehozására /ld. 2.3.2 pont/. Új adatok létrehozásakor a további lépésekben csak egyszerű összevonásokra kihagyásokra van lehetőség. Ezért néha a felhasználó egy szerteágazóbb, bonyolultabb feldolgozásra kényszerül, hiszen újra és újra vissza kell térni az első részrendszerhez.

A modulszerkezet továbbfejlesztésével lehetővé válik többváltozós függvények /összetett adatképzési eljárások/ létrehozása a rendszer minden fázisában.

3.2.1.2 Számítástechnikai feladatoknál, ha a felhasználó jól ismeri az alkalmazott rendszert, akkor esetenként sor kerülhet egyes részrendszerek változtatására is. Nagyon hasznos, hogy ha a módosítást elvégző program már eleve rendelkezésre áll. Ilyen program a SIS77 rendszerbe is beépíthető.

3.2.1.3 A SIS77 alkalmazásakor a teljes adatrendszer felbontásánál /2.3.3 pont/ a felhasználó többféle utat is követhet. Ugyanigy van a táblafájl-ok kialakításánál is /2.3.6 pont/. Elég nehéz azonban egy viszonylag jó hatásfokot biztosító feldolgozási folyamatot megtervezni. Kivánatos ezért egy olyan segédprogramot készíteni, amely adott információigény esetén meghatároz egy optimális adatrendszerfelbontást és táblafájl beosztást.

3.2.2 A rendszerműködés hatékonyságának növelése. Ez a kérdés elsősorban a rendszerműködtetés pénzügyi /géporadíj/ feltételeivel kapcsolatos, de akkor is előtérbe kerül, ha a rendelkezésre álló gépkapacitást túlhaladó feladatok megoldására van igény.

3.2.2.1 Az előzőkben /2.2.2 pont/ már említettük, hogy a pillanatnyi adattartalomhoz igazodva optimalizálni lehet egy adatfeldolgozó rendszer működését. A SIS77 ellenőrző, átkódoló részrendszerében már alkalmaztuk ezt az elvet /ld. [23], [36]/. A feldolgozás további lépéseiben - válogatások, táblafile-képzések - az egyes kódértékek eloszlásának ismeretében lehet hatékonyabbá tenni a feldolgozást /ld. [11]/.

Vizsgálataink szerint például a betegségek nemzetközi osztályozásában használt mintegy 3000-féle kódnak már 10 százaléka is elegendő az előforduló esetek közel 90 százalékának leírására. Így az esetek nagy többségénél nem szükséges a nehezen kezelhető négyjegyű kódot alkalmazni. /Ugyanez elmondható a betegségkódok összevont jegyzékeire is./

A feldolgozásmód kódérték szerinti szétválasztása /gyakori betegségek - ritka betegségek/, és az eredmények egyesítése a rendszer modulszerkezetének továbbfejlesztésével valósítható meg.

3.2.2.2 A táblafileokkal kapcsolatban az előzőkben /2.3.6.5 pont/ felvetett probléma, nevezetesen az, hogy a táblafileok terjedelme esetenként túlságosan megnőhet, az ellenőrző és átkódoló részrendszerben alkalmazott hierarchikus gráfstruktúra alkalmazásával megoldható. Ennek a lehetőségnek a kihasználása a SIS77 hatékonyságát jelentősen megnövelheti.

3.2.2.3 A SIS77 legidőigényesebb lépései az adatkonverziók. Ezért dolgoztunk ki egy olyan konvertálási eljárást /ld. [14], [34], [35] /, amely FORTRAN programokban egy a szokásosnál lényegesen gyorsabb konverziót tesz lehetővé. A kísérletben felhasznált HWB 66/60-as gépen 5-10 -szeres sebességnövekedést tapasztaltunk /ld. [14] /. Extrém esetekben ez az új konverziós módszer a hagyományos eljárással szemben közel százszoros sebességnövekedést is eredményezhet /ld. [34], [35] /.

Becsléseink szerint önmagában ez az egy sebességnövelési /konverziógyorsítási/ lehetőség is legalább 50 százalékos gépidőmegtakarítást eredményezhet a mostani kórházi morbiditási vizsgálatok feldolgozásánál. Természetesen ez az arány a megoldandó feladat függvényében lényegesen változhat. Egy ilyen mértékű sebességnövelés lehetővé teszi azt, hogy a SIS77 segítségével nagyobb adattömeg, például egy teljeskörű /100 százalékos/ kórházi morbiditási vizsgálat anyaga is könnyen feldolgozható legyen. Nagyméretű feladatoknál egyébként különösen jelentős egy adott arányú erőforrás /gépidő/ megtakarítás.

3.2.2.4 Az ellenőrző és átkódoló részrendszer /ld. 2.3.2 pont/ jelentős részét nem a SIS77-ben szokásos szerkesztési eljárások /ld. 2.2.2.3 pont/ hozzák létre, hanem előre rögzített utasításokból áll. A sebességnövelés

érdekében szükséges az, hogy ebben a részrendszerben is teljeskörűen alkalmazzuk a fent említett szerkesztő eljárásokat.

Ennek a részrendszernek egy másik hátránya az, hogy az adatellenőrzéseket és kódolásokat leíró hierarchikus gráfstrukturára /a biztonságos üzemeltetés érdekében/ kirótt szigorú feltételek miatt a kódolási, ellenőrzési eljárások leírásának megváltoztatása esetenként nehézkes és munkaigényes. Például a jelenlegi kórházi morbiditási vizsgálatban a kódolási és ellenőrzési eljárások /diagnózis, kor és nem összeférhetetlenségének vizsgálata, a foglalkozás és a kórház kódjának tömörítése, stb/ leírását - még a jelenlegi rendszer által biztosított igen tömör formában is - csak több mint ezer paraméterkártyán lehet megadni. Változás esetén ezt a paraméterrendszert újra összeállítani csak megfelelő képzettségű szakember tudja, kisebb módosításokat azonban alacsonyabb szinten /és sokkal kevesebb munkával/ is meg lehetne oldani. A kódolási szabályok felülvizsgálatával és módosításával lehet ezen a területen előbbre lépni.

Ugyancsak ennek a részrendszernek a továbbfejlesztése a memóriakihasználás optimalizálásának vonalán is lehetséges. Nagyméretű feladatoknál részben a program szegmentálásával /overlay/, részben az ellenőrzési és kódolási feladatok több lépésre való bontásával növelhető ennek a részrendszernek a hatékonysága.

I r o d a l o m j e g y z é k



- [1] AHO A.V. , DENNING P.J., ULLMANN J.D. , Principles of optimal page replacement, J. Assoc. Comp. Mach., 18., 1971.
- [2] ANSI/X3/SPARC Study Group, Interim Report, American National Standards Institute, 1975.
- [3] BENCZUR A., KRÁMLI A., PERGEL J. , On the Bayesian approach to optimal performance of page storage hierarchies, Acta Cybernetika, 3., 1977.
- [4] CODASYL Data Base Task Group Report, ACM, New York, 1971.
- [5] CODD E.G. , A Relational Model of Data for Large Shared Data Banks, CACM, 13., 1970.
- [6] COFFMAN E.G., KLEINROCK L. , Computer Scheduling Methods and their Countermeasures, Proceedings, AFIPS, 1968. SJCC.
- [7] CSÁKI P., GERGELY J., CZOPF J. , Kiváltott potenciálok szűrése nemstacionárius módszerrel, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 6. Kollokvium, Szeged, 1975.

- [8] CSERBA L. , Az általános kórházak betegforgalmi adatainak elemzése és felhasználásuk az egészségügyi tervezésben /1., 2./ , Egészségügyi Gazdasági Szemle, 3., 3-4., 1965.
- [9] CSUKÁS ANDRÁSNE, GREFF L., KRÁMLI A., RUDA M. , A kórházi morbiditási vizsgálat számítógépes feldolgozásának tapasztalatai és továbbfejlesztése, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 5. Kollokvium, Szeged, 1974.
- [10] CSUKÁS ANDRÁSNE, GREFF L., KRÁMLI A., RUDA M. , An approach to the hospital morbidity data system development in Hungary, Symposium on Medical Data Processing, Toulouse, 1975.
- [11] CSUKÁS ANDRÁSNE, GREFF L., KRÁMLI A., RUDA M. , Lekérdező rendszer kórházi morbiditás vizsgálat anyagára, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 6. Kollokvium, Szeged, 1975.
- [12] DATE C.J. , An Introduction to Database Systems, The systems programming series, Addison Wesley - Reading, Massachusetts - Menlo Park, California, London, Amsterdam - Don Mills, Ontario, Sydney /második kiadás/, 1976.

- [13] FINKELSTEIN M. , A Compiler Optimization Technique, Computer Journal, 2., 1968.
- [14] GÁL A., RUDA M. , Egy lehetőség Honeywell FORTRAN programok konverziós műveleteinek gyorsítására, SZÁMKI Tanulmányok, 1978.
- [15] GARÁDI J., RUDA M., KRÁMLI A., RATKÓ I. , Statisztikai és számítástechnikai módszerek alkalmazása kórházi morbiditási vizsgálatokban, MTA SZTAKI Tanulmányok, 35/1975.
- [16] H.GAUDI I., GYÁRFÁS I., CSUKÁS ANDRÁSNE , Az acut myocardialis infarctus fellépésének napi periodicitás vizsgálata, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 6. Kollokvium, Szeged, 1975.
- [17] GREFF L., KRÁMLI A., RUDA M. , Kórházi morbiditási vizsgálattal kapcsolatos statisztikai és számítástechnikai megfontolások, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 4. Kollokvium, Szeged, 1973.
- [18] GREFF L. , Hospitalizált morbiditás 1972-73 /1. kötet/, ESZTIK, Budapest, 1976.
- [19] GREFF L. , Hospitalizált morbiditás 1972-73 /2. kötet/, Egészségügyi Minisztérium, Budapest, 1974.

- [20] KNUTH D.E. , The art of computer programing, Sorting and searching /3. kötet/, Addison Wesley, London, California, 1973.
- [21] KRÁMLI A., RUDA M. , The computer realisation and first experiences of the hospital morbidity study, WHO Statisztikai Vándorszeminárium, Budapest, 1974. május 27.- június 3.
- [22] KRÁMLI A., RUDA M. , Izpravocsno-informacionnaja szisztjema zaproszov bolnyicsnovo morbigyityizma, Sztruktura i organyizacija paketov programm, Nemzetközi Konferencia, Tbiliszi, 1976.
- [23] KRÁMLI A., RATKÓ I., RUDA M., SOLTÉSZ J. , A statisztikai adatfeldolgozás matematikai és számítástechnikai problémái, MTA SZTAKI Tanulmányok, 70/1977.
- [24] MARTIN J. , Computer Data-Base Organization, Prentice-Hall, New Jersey, 1975.
- [25] MÁNYI G. , A hospitalizált morbiditás strukturája Magyarországon, ESZTIK, Budapest, 1973.
- [26] MTA SZTAKI dokumentáció, Az 1972-73. évi kórházi morbiditási vizsgálat számítógépes feldolgozása /1. és 2. kötet/, Budapest, 1974.
- [27] NIEVERGELT J. , On the Automatic Simplification of Computer Programs, CACM, 8., 1965.

- [28] RATKÓ I., CSUKÁS ANDRÁSNÉ, JÁNOSI A., GYÁRFÁS I. ,
Az Infarctus Regiszter működése számítógépes rendszerrel, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 8. Kollokvium, Szeged, 1977.
- [29] RATKÓ I. , Egy számítástechnikai eszköz bonyolult logikai kifejezések leírására orvosstatisztikai alkalmazásokban, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 8. Kollokvium, Szeged, 1977.
- [30] RATKÓ I., Bonyolult logikai kifejezések kiértékelésének számítástechnikai és optimalizálási problémái, MTA SZTAKI Közlemények, 20/1978.
- [31] RUDA M., SZEIDL L. , Idősorelemzés I., MTA SZTAKI, Budapest, 1973.
- [32] RUDA M. , Egy általános információs rendszer korhízi morbiditási adatok feldolgozására, Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és a biológiában, 8. Kollokvium, Szeged, 1977.
- [33] RUDA M. , Statistical Information System with Health Service Application, 4. Winterschool of Visegrád on the Theory of Operating System, Szentendre, 1978.

- [34] RUDA M. , Egy széles körben alkalmazható programoptimalizálási módszer, MTA SZTAKI Közlemények, 20/1978.
- [35] RUDA M. , Egy módszer programok erőforrásigényének csökkentésére és bonyolult szerkezetű programok egyszerűsítésére, Számítástechnika, IX. évf., 1978.
- [36] RUDA M. , Egy számítástechnikai módszer függvénytáblázatok tömör tárolására, egy adatfeldolgozási alkalmazással /kézirat/.
- [37] RUDA M., SZEIDL L., TUSNÁDY G. , Time series analysis on CDC 3300, 9. European Meeting of Statisticians, Budapest, 1972.

